TO:Auto-Reply fax 352 372 5800 COMPANY:

Auto-Reply Facsimile Transmission

UNITED STATES PATENT AND TRADEMARK OFFICE

TO:	Fax Sender at 352 372 5800
Fax Information	
Date Received:	3/25/02 2:46:32 PM [Eastern Standard Time]
Total Pages:	5 (including cover page)

ADVISORY: This is an automatically generated return receipt confirmation of the facsimile transmission received by the Office. Please check to make sure that the number of pages listed as received in Total Pages above matches what was intended to be sent. Applicants are advised to retain this receipt in the unlikely event that proof of this facsimile transmission is necessary. Applicants are also advised to use the certificate of facsimile transmission procedures set forth in 37 CFR 1.8(a) and (b), 37 CFR 1.6(f). Trademark Applicants, also see the Trademark Manual of Examining Procedure (TMEP) section 702.04 et seq.

Received Cover Page ========>

MAR-25-2002 MON 02:41 PM SALIWANCHIK, LLOYD&SALIWA FAX NO. 352 372 5800 P. 01

FACSIMILE COVER SHEET

SALIWANCHIK, LLOYD & SALIWANCHIK
A Professional Association
2421 N.W. 41st Street
Suite A-1
Gainesville, FL 32606

Telephone (352) 375-8100
Facsimile (352) 372-5800

TO: Examiner Jeffrey Norman Fredman **FROM:** Frank C. Eisenschenk, Ph.D.

COMPANY: U.S. Patent Office, Art Unit 1655 **DATE:** March 25, 2002

FAX NO.: 1-703-872-9306 **NO. OF PAGES: (INCLUDING COVER SHEET)** 5

SUBJECT/MESSAGE: U.S. Patent Application Docket No. GEN-T124X
GENES, PROTEINS AND BIALLELIC MARKERS RELATED OT CENTRAL NERVOUS SYSTEM DISEASE
(Blumenfeld, *et al.*)
Serial No. 09/416,384; Filed October 12, 1999

THIS IS AN OFFICIAL DOCUMENT.

PLEASE DELIVER IMMEDIATELY.

Attachments: Supplemental Amendment Under 37 C.F.R. § 1.111
Marked-Up Version of Amended Claims

If you do not receive all pages or if any transmission is not legible, call the sender at (352) 375-8100.

H:\FCE\Gen\T124X\Fax 2Examiner.doc\DNB/jaj

Received from <352 372 5800> at 3/25/02 2:46:32 PM [Eastern Standard Time]

DOCKET NO.: GEN-T124X February 19, 2002
APPLICANTS: Marta Blumenfeld, et al.
SERIAL NO.: 09/416,384
FILED ON : October 12, 1999

SUBMISSION TO PTO:

1. Amendment Under 37 C.F.R. 1.111
2. Exhibits A-F as listed on the last page of the Amendment

jaj

60/4

I hereby certify that this correspondence is being deposited with the United States Postal Service as First Class Mail in an envelope addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231 on February 19, 2002.

Frank C. Eisenschenk

Frank C. Eisenschenk, Ph.D., Patent Attorney

AMENDMENT UNDER 37 C.F.R. § 1.111
Examining Group 1655
Patent Application
Docket No. GEN-T124X
Serial No. 09/416,384

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Examiner	:	Jeffrey Fredman, Ph.D.
Art Unit	:	1655
Applicants	:	Marta Blumenfeld, Lydie Bougueleret, Ilya Chumakov, Daniel Cohen, Laurent Essioux
Serial No.	:	09/416,384
Filed	:	October 12, 1999
Confirm. No.	:	6101
For	:	GENES, PROTEINS AND BIALLELIC MARKERS RELATED TO CENTRAL NERVOUS SYSTEM DISEASE

Assistant Commissioner for Patents
Washington, D.C. 20231

RESPONSE AND AMENDMENT UNDER 37 C.F.R. § 1.111

Remarks

Claims 58, 62, and 73-75 are pending in the subject application. The Office Action of November 19, 2001 has set forth a variety of rejections

Rejection under 35 U.S.C. § 101:

The Office Action, dated November 19, 2001, has rejected claims 58, 62, and 73-75 under 35 U.S.C. § 101 for allegedly lacking utility. Specifically, the Office Action alleges that the only taught utilities of the protein (encompassed by the present claims) are to detect the protein, to make

antibodies, and to screen drugs, and that while these utilities are credible, they are allegedly neither specific nor substantial (pages 2-5 of the November 19, 2001 Office Action). Applicants respectfully traverse this rejection.

Applicants maintain that the utilities the Patent Office acknowledges to be credible in the Office Action are also specific and substantial. In particular, in view of: a) the strong association between the genomic region encoding the present protein and schizophrenia; b) the fact that the present protein is expressed mainly in the brain; c) the fact that the protein is a novel member of a family of proteins frequently implicated in neuropsychiatric disorders; and d) G713 being a candidate gene for neuropsychiatric disorders, on the basis of both structure and function (see *infra*); Applicants submit that the present protein has specific and substantial utility in the identification of individuals at risk for the development of neurodegenerative disorders or the treatment and diagnosis of neuropsychiatric disorders such as schizophrenia.

Substantial utility

The Office Action alleges that the cited utilities of the protein encompassed by the present claims that are provided in the specification are not substantial. Applicants respectfully traverse.

In particular, the Office Action states that evidence of the association of all glutamine repeat proteins with any disease state might be sufficient to provide a substantial utility. Applicants respectfully submit that this is not the standard for the establishment of the utility of an invention. The evidentiary standard for application in this matter is "preponderance of the evidence". Thus, the issue is more properly framed as, "Is it more likely than unlikely that proteins containing polyglutamine repeats are associated with neurodegenerative diseases?" To this end, the Office Action indicates that:

"...many [genes] have glutamine repeats. A search of STN revealed that 14,560 sequences in the database which had a 3 repeat sequence of CAGCAGCAG, of which 6,723 were associated with human sequences" (page 4, November 19, 2001 Office Action).

Applicants respectfully submit that this analysis is irrelevant to the question at hand. As recognized by the Patent Office, previously presented in arguments by Applicants' counsel, and presented herein, proteins containing glutamine repeats, and expansions of such repeats, are associated with at least twelve (12) diseases, including neurodegenerative diseases. Thus, Applicants respectfully submit that those skilled in the art are well aware of the relationship between glutamine repeats and neurological disorders. Furthermore, Applicants respectfully submit that, on the basis of its expression in brain and its CAG (glutamine) repeat (among other hallmarks of repeats known to undergo expansion in disease states), a person of ordinary skill in the art would "consider [G713] a candidate gene for neuropsychiatric disorder on the basis of both structure and function" [paragraph bridging pages 119-20; Margolis *et al.* [1997] *Human Genetics* 100:114-122; Exhibit A]. Additionally, given the occurrence of sequential repeats of four glutamines, nine glutamines, and six glutamines within the G713 protein, Applicants respectfully submit that a person of ordinary skill in the art would perceive the use, by the Patent Office, of "a 3 repeat sequence of CAGCAGCAG [$(CAG)_3$]" rather than a 9-19 repeat sequence of $(CAG)_{9-19}$ to be inappropriate for assessing the significance of the glutamine repeat within G713.

It is further respectfully submitted that an absolute number of 6,723 hits for proteins containing a three repeat glutamine sequence is meaningless. In the context of this invention, it is the number and relationship of glutamine repeats with neurodegenerative disorders that are of

significance. Applicants respectfully submit that a more meaningful number is obtained from the paper by Margolis *et al.* (Exhibit A), wherein it was found that about 20 distinct cDNA clones were isolated from screening 200,000 or more plaques of human brain cDNA libraries with $(CAG)_{15}$. This corresponds to a representation of glutamine repeat within brain cDNA on the order of about 0.01%, a value inconsistent with the Patent Office's assertion that "*many* [genes] have glutamine repeats" [emphasis added; page 4, November 19, 2001 Office Action]. Indeed, the skilled artisan would reasonably expect that genes and proteins containing high numbers of glutamine repeats are associated with neurodegenerative disorders on the basis of structure and function in neurological tissue (Margolis *et al.*, Exhibit A, paragraph bridging pages 119-120). Thus, Applicants respectfully submit that a person of ordinary skill in the art would find reasonable Applicants' conclusion that G713 is suitable for screening as a marker for neuropsychiatric disorders on the basis of both structure and function.

Applicants also respectfully point out that the M.P.E.P. at § 2107.02 states, with reference to *In re Langer*:

> As a matter of Patent Office practice, a specification which contains a disclosure of utility which corresponds in scope to the subject matter sought to be patented must be taken as sufficient to satisfy the utility requirement of § 101 for the entire claimed subject matter unless there is a reason for one skilled in the art to question the objective truth of the statement of utility or its scope.

Applicants respectfully submit that: 1) in view of the strong association between the genomic region encoding the present protein and schizophrenia; 2) in view of the fact that the present protein is expressed mainly in the brain; 3) in view of the fact that the protein is a novel member of a family of proteins frequently implicated in neuropsychiatric disorders; and 4) in view of G713 being a candidate gene for neuropsychiatric disorder on the basis of both structure and function, that there is

a correlation or association of the present protein with predisposition to the onset of neuropsychiatric disorders such as schizophrenia. As stated in the M.P.E.P. at § 2107.01:

> *An assay that measures the presence of a material which has a stated correlation to a predisposition to the onset of a particular disease condition* [emphasis added] would also define a "real world" context of use in identifying potential candidates for preventive measures or further monitoring.

As one of the cited utilities of the protein encompassed by the present claims that are provided in the specification is to detect the protein (see *supra*), Applicants respectfully submit that the utilities, of the protein encompassed by the present claims, provided in the specification are substantial.

Specific utility

The Office Action also alleges that cited utilities, of the protein encompassed by the present claims, that are provided in the specification are "not specific because Perutz ([1996] *Current Opinion Structural Biology* 6:848-58) ... has identified many different proteins with glutamine repeats, all of which are associated with different diseases and Kashima has identified at least one protein not disease associated with glutamine repeats" (page 4, November 19, 2001 Office Action).

Applicants respectfully dispute the Patent Office's representation of the Perutz paper. In the first sentence, Perutz states: "*Several* [emphasis added] dominantly inherited, late onset, *neurodegenerative diseases* [emphasis added] are due to expansion of CAG repeats, leading to expansion of glutamine repeats in the affected proteins." Perutz then proceeds to discuss only *seven* neurodegenerative diseases. Furthermore, as indicated in Margolis *et al.*, twelve (12) diseases, most with neurotrophic features, arise from trinucleotide repeat expansion mutations (see Abstract and Introduction).

Applicants further respectfully traverse the relevance of the Kashima paper. The observation by Kashima *et al.* that "the consecutive glutamine repeats do not play a role in the biological and immunological activities of MIL-2" (page 4, November 19, 2001 Office Action) is irrelevant to examination of the present application. Applicants further respectfully submit that as MIL-2 is not expressed in brain, MIL-2 is *a priori* irrelevant to any discussion of the role of CAG repeats, and their expansion, in neuropsychiatric disorders such as schizophrenia. Furthermore, Kashima *et al.*, and the teachings therein, are directed to the structure and function of murine interleukin 2 (MIL-2) and contain no teachings with respect to neurodegenerative disorders. Thus, it is unclear what, if any, nexus or evidentiary value the reference has to the assessment of the role of glutamine repeats in neurodegenerative disorders. Furthermore, as the Office Action acknowledges, glutamine repeats, and their expansion, are art recognized to be associated with a variety of neurodegenerative disorders, such as schizophrenia. Accordingly, reconsideration and withdrawal of the rejection is respectfully requested.

Rejection under 35 U.S.C. § 112:

In the Office Action dated November 19, 2001, the Patent Office rejected claims 58, 62, and 73-75 under 35 U.S.C. § 112, first paragraph, as allegedly containing subject matter which was not described in the specification in such a way as to enable one skilled in the art to which it pertains, or with which it is most nearly connected, to use the invention. Specifically, the Office Action alleges that the cited utilities, of the protein encompassed by the present claims, that are provided in the specification allegedly are not specific. Applicants respectfully traverse this rejection.

Nature of Invention

Claims are drawn to a G713 protein and methods of detection of this protein or gene product. The Office Action alleges that "the nature of this invention is a ... protein, with no other associated information" [page 5-6, November 19, 2001 Office Action].

Applicants respectfully traverse this representation of the invention in view of: 1) the strong association between the genomic region encoding the present protein and schizophrenia; 2) the fact that the present protein is expressed mainly in the brain; 3) the fact that the protein is a novel member of a family of proteins frequently implicated in neuropsychiatric disorders; and 4) the fact that G713 is a candidate gene for neuropsychiatric disorder on the basis of both structure and function (see *supra* and *infra*).

Breadth of the Claims

As the Patent Office acknowledges, "claims ... are drawn to a G713 protein and methods of detection of this protein or gene product" [page 5, November 19, 2001 Office Action].

Amount of Guidance in the Specification

The Office Action alleges that the utilities of the protein encompassed by the present claims that are provided in the specification are "not found to be substantial nor specific and, consequently, the specification provides NO guidance regarding how to use this protein" [page 6, November 19, 2001 Office Action]. Applicants respectfully disagree. As indicated *supra,* the subject invention provides credible, substantial, and specific uses of the presently claimed invention. Applicants submit that the argument of a consequential lack of "guidance regarding how to use this protein", arising from the allegation that the invention lacks specific and substantial utility, is improper and respectfully request reconsideration and withdrawal of this point of rejection.

Working Examples

The Office Action also alleges that "the absence of working examples in the specification" [page 8, November 19, 2001 Office Action] is a valid basis for a conclusion that "undue experimentation would be required to use this invention as claimed" [page 8, November 19, 2001 Office Action]. Applicants respectfully traverse this point of rejection and submit that a person of ordinary skill in the art would know how to use the present invention in an "assay for detection or diagnosis" [page 6, November 19, 2001 Office Action]. In support of Applicants' position, it is noted that just one type of assay encompassed by the method of claim 73, namely enzyme linked immunosorbent assays, is associated with at least 3307 references in PubMed [Exhibit B].

Amount of Guidance in Prior Art

The Office Action also argues that "the prior art provides no guidance with regard to the particular function of the G713 protein and does not even provide support or guidance for glutamine repeat containing proteins having a particular use or association" [page 6, November 19, 2001 Office Action]. It is respectfully submitted that this allegation is without basis in view of the art-recognized association of glutamine repeats with neurodegenerative diseases (see, for example, Margolis *et al.*).

The Patent Office also reiterates an argument based on the Kashima paper. As indicated supra, the relevance and evidentiary value of the Kashima paper to the instant invention is unclear. The reference concerns the role of glutamine repeats in the biological function of a murine cytokine and is devoid of any teachings related to glutamine repeats and their association with neurodegenerative disorders. Applicants further reiterate that in the context of the present invention, whether glutamine repeats play a role in the normal function of a given protein is irrelevant; as

explicitly acknowledged elsewhere by the Patent Office, high numbers of glutamine repeats, and their expansion, are hallmarks of neurodegenerative disorders and the proteins associated therewith.

Skill in the Art

The Patent Office believes that "the skill in the art would be considered high" [page 7, November 19, 2001 Office Action]. Applicants concur.

Predictability of the Art

The Patent Office cites a review article by Wright *et al.* ([2001] *Schizophrenia Research* 47:1-12) to support his assertion of the "unpredictability in this linkage of schizophrenia with chromosomal locations" (page 7, November 19, 2001 Office Action). Applicants respectfully dispute the Patent Office's representation of linkage studies of schizophrenia as unpredictable. Genetic analysis of schizophrenia and other diseases believed to have a number of susceptibility loci are complicated but *not invalidated* by issues of population stratification and statistical power. These considerations are well appreciated by persons of ordinary skill in the art and are discussed, for example, in Weinberger *et al.* ([2001] *Biological Psychiatry* 50:825-44; Exhibit C).

The issue of population stratification is well known to Applicants as evidenced by the analysis presented in Example 2(g) of the specification (pages 164, 167, and 174). Example 2(g) presents the results of an association study between schizophrenia and the biallelic markers of the invention. In that study, it was found that "the relative difference in Mean normalized LD for BAC B5 is significantly higher when the comparison was made between *familial cases* and controls than when the comparison was made between the *whole cases* and the controls [emphasis added]" (page 167).

That the issue of population stratification is well known to persons of ordinary skill in the art is evidenced by the study presented in the paper by Brant *et al.* ([2000] *Gastroenterology* 119:1483-90; Exhibit D) dealing with genetic analysis of the analogously complex Crohn's disease (CD) and as discussed therein with reference to other studies. In the paper by Brant *et al.*, it was found that: "Pedigrees with CD diagnosed at an earlier age have greater linkage evidence for IBD1. This effect is greatly magnified when the *individuals with younger onset also have relatively more severe disease*..." (page 1488) [emphasis added].

The issue of statistical power is well known to Applicants as evidenced by patents (for example, U.S. Patent No. 6,291,182; Exhibit E) and proprietary software directed to the statistical analysis involved in determining whether a genomic region is associated with a detectable trait such as schizophrenia.

Applicants further dispute the Patent Office's assertion of "unpredictability in this linking of schizophrenia with chromosomal locations" (page 7, November 19, 2001 Office Action) in view of Applicants' success in further associating schizophrenia with gene g35030 (WO 01/40493; Exhibit F) and in view of others' success in identifying an association of catechol-o-methyl transferase (COMT) gene with schizophrenia [Weinberger *et al.* [2001] *Biological Psychiatry* 50:825-44; Exhibit C].

In view of the foregoing evidence, Applicants respectfully submit that the Patent Office allegation regarding the "unpredictability in this linkage of schizophrenia with chromosomal locations" (page 7, November 19, 2001 Office Action) is improper. Reconsideration and withdrawal of the rejection is respectfully requested.

Quantity of Experimentation

The Office Action alleges that "an immense amount of experimentation would be required in order to define whether this protein is associated with any particular disease state" and concludes that "undue experimentation would be required to use this invention as claimed" (page 7, November 19, 2001 Office Action). Applicants respectfully dispute this representation of the invention in view of the strong association between the genomic region encoding the present protein and schizophrenia, in view of the fact that the present protein is expressed mainly in the brain, in view of the fact that the protein is a novel member of a family of proteins frequently implicated in neuropsychiatric disorders, and in view of G713 structure and function that is suggestive of roles in neurodegenerative disorders. It is further submitted that that there is a correlation of the present protein to a predisposition to the onset of neuropsychiatric disorders such as schizophrenia (see *supra*). Applicants respectfully submit that the Patent Office has failed to establish a *prima facie* case regarding a lack of enablement and respectfully request reconsideration and withdrawal of the rejection.

In view of the foregoing remarks and amendments to the claims, Applicants believe that the currently pending claims are in condition for allowance, and such action is respectfully requested.

The Commissioner is hereby authorized to charge any fees under 37 C.F.R. §§ 1.16 or 1.17 as required by this paper to Deposit Account No. 19-0065.

Applicants invite the Examiner to call the undersigned if clarification is needed on any of this response, or if the Examiner believes a telephonic interview would expedite the prosecution of the subject application to completion.

Respectfully submitted,

Frank C. Eisenschenk, Ph.D.
Patent Attorney
Registration No. 45,332
Phone No.: 352-375-8100
Fax No.: 352-372-5800
Address: Saliwanchik, Lloyd & Saliwanchik
A Professional Association
2421 NW 41st Street, Suite A-1
Gainesville, FL 32606-6669

FCE/jaj

Attachments:
Exhibit A: Margolis *et al.* [1997] *Human Genetics* 100:114-122
Exhibit B: List of 3307 references in PubMed
Exhibit C: Weinberger *et al.* [2001] *Biological Psychiatry* 50:825-44
Exhibit D: Brant *et al.* [2000] *Gastroenterology* 119:1483-90
Exhibit E: U.S. Patent No. 6,291,182
Exhibit F: Patent No. WO 01/40493

Am. J. Hum. Genet. 54:864-870, 1994

Evidence for Anticipation in Schizophrenia

Anne S. Bassett* and William G. Honer†

*Department of Psychiatry, University of Toronto, and Genetics Section, Schizophrenia Research Program, Queen Street Mental Health Centre, Toronto; and †Department of Psychiatry, University of British Columbia, Vancouver

Summary

Anticipation, or increasing severity of a disorder across successive generations, is a genetic phenomenon with an identified molecular mechanism: expansion of unstable trinucleotide repeat sequences. This study examined anticipation in familial schizophrenia. Three generations of siblines from the affected side of families selected for unilineal, autosomal dominant-like inheritance of schizophrenia were studied (n = 186). Across generations more subjects were hospitalized with psychotic illness ($P<.0001$), at progressively earlier ages ($P<.0001$), and with increasing severity of illness ($P<.0003$). The results indicate that anticipation is present in familial schizophrenia. These findings support both an active search for unstable trinucleotide repeat sequences in schizophrenia and reconsideration of the genetic model used for linkage studies in this disorder.

Introduction

The clinical observation of anticipation—i.e., inherited illness that becomes more severe across successive generations—has recently been found to have a molecular basis: expanding GC-rich trinucleotide repeat sequence mutations (Harper et al. 1992; Sutherland and Richards 1992). In fragile X syndrome (Verkerk et al. 1991), myotonic dystrophy (Fu et al. 1992), spinobulbar muscular atrophy (Brook et al. 1992), spinocerebellar atrophy type 1 (Orr et al. 1993), and Huntington disease (Huntington's Disease Collaborative Research Group 1993), increasing severity of illness, earlier age at onset, and/or increasing proportion of ill individuals in successive generations are associated with longer trinucleotide expansions. Schizophrenia is another neuropsychiatric disorder that may display this anticipation phenomenon and that therefore may have familial forms caused by an unstable trinucleotide repeat.

Schizophrenia is a severe disorder characterized by social withdrawal and psychotic symptoms, such as delusions and hallucinations. The illness has a variable age at onset, often beginning in early adulthood and resulting in lifelong disabilities in social and occupational functioning. Evidence from family, twin, and adoption studies, including those using reliable diagnostic criteria (Lowing et al. 1983; Kendler et al. 1985), strongly supports a genetic etiology for schizophrenia (Gottesman and Shields 1982). However, the mode of inheritance for schizophrenia is not readily identifiable and is proposed to involve interacting genes (Risch 1990). In families with the illness, reduced penetrance and variable expression are commonly found. Other psychotic disorders, of lesser severity, and schizotypal personality traits such as social isolation, odd communication, and extreme suspiciousness are conditions likely reflecting variable expression of genetic susceptibility to schizophrenia (Gottesman and Shields 1982; Lowing et al. 1983; Kendler et al. 1985). These factors, along with the possibility of genetic heterogeneity and the practical difficulties of studying an illness with a significant suicide rate and suspicious, socially isolated individuals, combine to make schizophrenia a challenging disorder for linkage studies (Bassett 1991). Strategies to overcome these difficulties include focusing on familial schizophrenia where inheritance is consistent with Mendelian patterns, using reliable diagnostic methods, highly polymorphic DNA markers, and lod-score methods that model the complexities of the inheritance. Linkage studies to date, using informative fami-

Received September 13, 1993; accepted for publication December 22, 1993.

Address for correspondence and reprints: Dr. Anne S. Bassett, Schizophrenia Research Program, Queen Street Mental Health Centre, 1001 Queen Street West, Toronto, Ontario M6J 1H4, Canada.

0002-9297/94/5405-0016$02.00

lies and models based on Mendelian inheritance, have yielded significant lod scores but no replicated positive results (Bassett 1991).

Dynamic modifications of classical patterns of genetic transmission, such as anticipation (Mott 1911) and genomic imprinting (differential expression of genetic material depending on parental origin of the gene and the underlying molecular mechanisms), may explain the complex genetics of schizophrenia and other major mental illnesses (McInnis et al. 1991; Flint 1992). For example, trinucleotide repeats can cause reduced penetrance and variable expression (Caskey et al. 1992; Sutherland and Richards 1992), by existing in a premutation form, by reductions in repeat size, or by somatic mutation in early embryogenesis (Lavedan et al. 1993). The unstable nature of trinucleotide repeats also provides a possible mechanism for the high mutation rates proposed for schizophrenia (Slater and Cowie 1971). If there were evidence of anticipation in schizophrenia, screening for triplet repeat mutations would become a rational option for gene localization studies, and modification of the genetic model used in linkage studies would need to be considered. The current study investigated whether anticipation was present in a familial schizophrenia sample participating in a linkage study.

Subjects and Methods

Subjects were members of eight extended nonconsanguineous families participating in a genetic linkage study of familial schizophrenia. Local psychiatrists identified prospective pedigrees segregating schizophrenia. Families were selected for large size, availability of two or more generations of adults, and apparent unilineal, autosomal dominant-like inheritance of schizophrenia and genetically related disorders. Bilineal families with evidence, from family or collateral history, of schizophrenia or other nonaffective psychotic disorders on both sides were excluded. Further details of the original ascertainment and assessment for the linkage study are described elsewhere (Bassett et al. 1993). Since families were ascertained in their entirety, proband status could be assigned to all affected subjects. Therefore no subjects were excluded from analyses.

Subjects only from the affected side of each family were taken into account, to determine sibling sets (siblines) in the index generation (IG) (n = 13 siblines), parental generation (PG) (n = 10 siblines), and grandparental generation (GG) (n = 8 siblines). The affected side was defined by (1) a parent hospitalized with psychosis (four cases), (2) an aunt/uncle hospitalized with psychosis (four cases), (3) a parent/aunt/uncle with schizotypal traits (seven cases), or (4) a sibship linking two affected nuclear families (six cases in two extended families) (see fig. 1). In two cases in the GG, the affected side could not be determined, and the smaller of the unaffected maternal or paternal siblines was arbitrarily selected. There were three instances of unknown paternity. In two of these cases, the maternal line was affected. In the third case, neither the mother nor her five siblings were affected, and this GG sibline was not included in the analysis.

Family-history information was obtained for each subject from three or more family members by using the Family History–Research Diagnostic Criteria (FH-RDC) method (Andreasen et al. 1977). Genealogical records were used to confirm dates of birth and death. Medical records were searched back to 1866 for evidence of psychiatric hospitalization. Because the subject families originated and seldom moved from a circumscribed region of Canada, and because the one psychiatric hospital available until the 1980s consistently maintained a comprehensive file-card system of recording admissions, virtually complete ascertainment of psychiatric hospitalization was assured. Records were collected for all subjects with a history of psychiatric hospitalization. Living subjects participating in the linkage study were directly interviewed by a psychiatrist (Bassett et al. 1993). Diagnostic folders containing the family history and, if present, medical records and interview data were reviewed independently by two psychiatrists (A.S.B. and W.G.H.), one of whom (W.G.H.) was blind to the pedigree structures. A consensus lifetime Research Diagnostic Criteria (RDC) diagnosis for psychotic disorders, age at first hospitalization for a psychotic illness, and presence of two or more RDC schizotypal traits were recorded. Psychotic disorders included schizophrenia (n = 25), schizoaffective disorder (n = 13; 12 mainly schizophrenic type and 1 other), unspecified functional psychosis (n = 4), mania with psychosis (n = 1), and depression with psychosis (n = 1). Schizophrenia and schizoaffective disorder were approximately equivalent in severity in the current sample (Bassett et al. 1993) and were considered together in the current study. Individuals with psychotic disorders not severe enough to require hospitalization, as well as subjects with schizotypal traits, were combined in a single schizotypal group, because of small numbers in each category.

Of the 209 subjects in the affected siblines, 23 were excluded from the analyses. One IG subject had not attained the age of 15 years, considered a minimum age

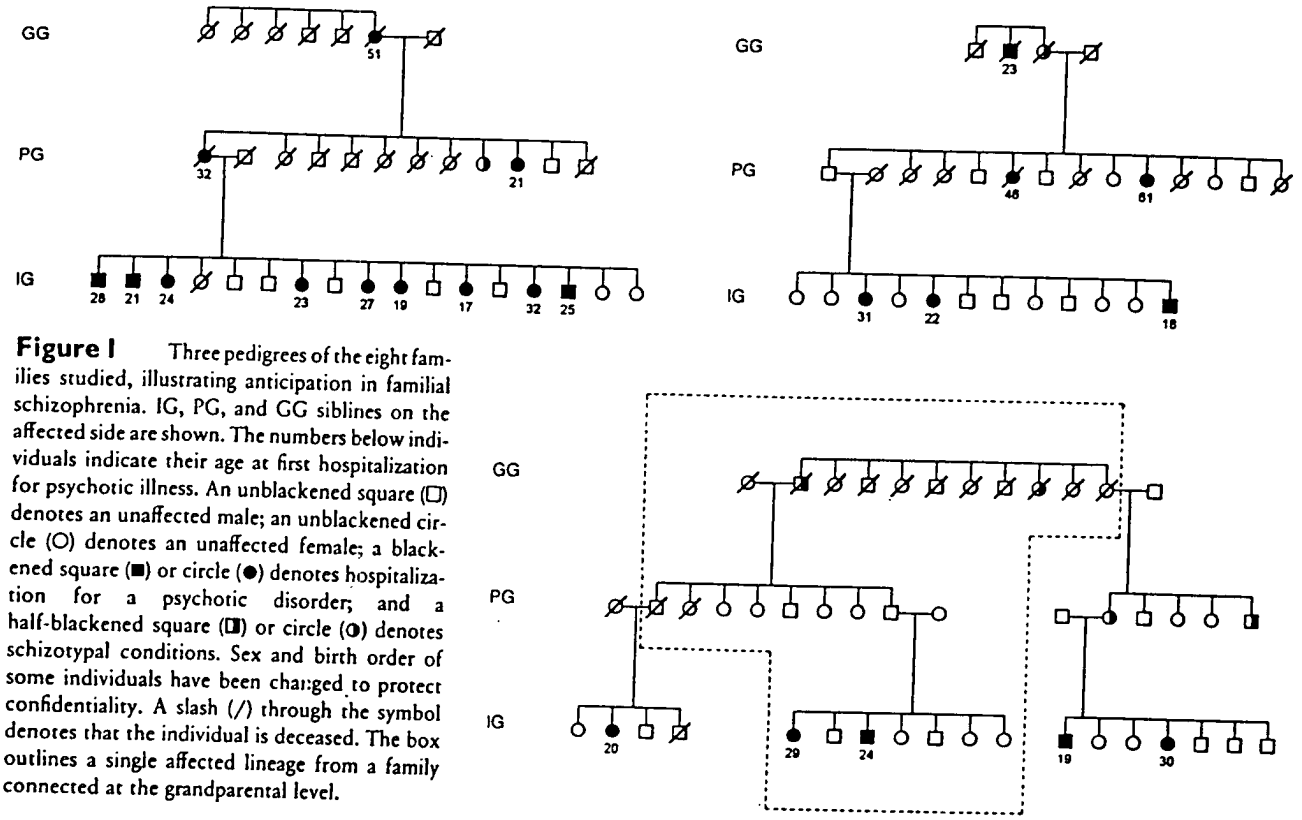


Figure 1 Three pedigrees of the eight families studied, illustrating anticipation in familial schizophrenia. IG, PG, and GG siblines on the affected side are shown. The numbers below individuals indicate their age at first hospitalization for psychotic illness. An unblackened square (□) denotes an unaffected male; an unblackened circle (○) denotes an unaffected female; a blackened square (■) or circle (●) denotes hospitalization for a psychotic disorder; and a half-blackened square (◨) or circle (◑) denotes schizotypal conditions. Sex and birth order of some individuals have been changed to protect confidentiality. A slash (/) through the symbol denotes that the individual is deceased. The box outlines a single affected lineage from a family connected at the grandparental level.

of risk for psychotic illness (Gottesman and Shields 1982). Five subjects (2 PG and 3 GG) had moved, and collateral information was insufficient to determine hospitalization status; and 17 subjects (3 IG, 5 PG, and 9 GG) died before the age of 40 years. Thirteen died in infancy or childhood, two in war, one in an accident at work, and one of unknown cause; none were suicides. Most new cases of schizophrenia may be expected before age 40 years (Gottesman and Shields 1982). Data on the remaining 186 subjects were examined for anticipation, in three ways. First, the rates of hospitalization for psychotic disorders and the rates of schizotypal conditions were compared across generations by using χ^2 analyses. Second, to assess severity of illness, subjects were assigned the following ratings: hospitalized with schizophrenia or schizoaffective disorders—3; hospitalized with other psychotic disorders—2; schizotypal—1; and unaffected—0. Means for each generation were compared using the one-way analysis of variance (ANOVA), including correction for multiple tests of significance with the Student-Newman-Keuls procedure. Third, age at first hospitalization for psychosis was assessed using the life-table method of survival analysis for 1-year intervals. Homogeneity of survival curves over the generations was examined using the Wilcoxon test. The analysis was performed assuming (1) no differential mortality between affected and unaffected and (2) hospitalization rates independent of chronological time (e.g., 1920 vs. 1970). Observations ended at the subject's current age or age at death. Covariates tested were sex and transmission patterns (maternal/paternal).

Results

Demographic characteristics of the sample and results indicating anticipation are presented in table 1. As for other illnesses demonstrating anticipation, expression of illness varied between members of a sibship (fig. 1).

Rates and Distribution of Illness

There were significantly more subjects hospitalized for psychosis across successive generations ($\chi^2 = 16.84$, $P<.0001$, 2 df). Most had schizophrenia or schizoaffective disorders (IG, $n = 30$; PG, $n = 8$; and GG, $n = 0$). Of the six subjects with less severe disorders—unspecified functional psychosis ($n = 4$), psychotic mania ($n = 1$), or psychotic depression ($n = 1$)—four were in PG or GG. Subjects with the least severe illnesses (schizotypal con-

Table 1

Characteristics of the Sample, by Generation (n = 186)

	IG	PG	GG
Total no. of subjects (females)	86 (38)	62 (24)	38 (21)
Mean size of sibline (SD)	6.61 (4.27)	6.20 (3.61)	5.43 (4.20)
Mean age of living[a] (SD)	40.61 (8.74)	65.55 (12.96)	71.80 (8.26)
Mean age at death[b] (SD)	44.25 (3.20)	57.15 (14.58)	77.35 (9.34)
Mean age at first hospitalization (SD)	26.16 (8.28)	34.00 (17.28)	41.33 (15.89)
Subjects hospitalized for psychotic illness	32 (37.21%)	9 (14.52%)	3 (7.89%)
Subjects with schizotypal conditions	8 (9.30%)	8 (12.90%)	8 (21.05%)

[a] No. of living subjects: IG, 82; PG, 42; and GG, 10.

[b] No. of dead subjects with known age at death: IG, 4; PG, 20; and GG, 20. Eight other GG subjects had ages at death that were less precisely known, e.g., "in their 70s." The mean shown did not change significantly when these subjects were included using estimated ages at death.

ditions) were twice as common in the GG as in the IG (21.05% vs. 9.30%); however, across the three generations, the result was not significant (χ^2 = 3.24, P = .20, 2 df). When these subjects with schizotypal conditions were included with hospitalized subjects, there were still significantly more affected subjects in the youngest generation (χ^2 = 6.86, P = .032, 2 df).

Because of the possibility of compounded error with families being connected at the parental or grandparental level, analyses were rerun with a single affected lineage from each of the seven kindreds that had data from all three generations, with the largest siblines being selected (see fig. 1). The same results were found using this subsample of 151 subjects, for both hospitalization rates (χ^2 = 12.30, P<.002, 2 df) and hospitalization plus schizotypal rates (χ^2 = 9.46, P<.009, 2 df). Excluding subjects who died before age 40 or moved away (n = 22) could have influenced the study's findings, since most were in the two senior generations. When results with these subjects and with penetrance estimated maximally at 100% were considered, so that one-half (n = 11) would have been hospitalized with psychotic disorders, the results for hospitalization rates would still have remained significant (χ^2 = 10.02, P<.01, 2 df). Data were also reanalyzed examining the effect that increasing hospitalization rates over time would have on the results. Secular trends of hospitalization rates for specific illnesses were not available. Therefore, arbitrary increases were tested, by adding 100% more subjects (n = 3) to the hospitalized GG group and 50% more (n = 4.5) to the PG group. Under these conditions, a significant increase in the rate of psychosis requiring hospitalization would continue to be present over the generations (χ^2 = 7.75, P<.05, 2 df).

Severity of Illness

Means for four-point severity-of-illness ratings for the three generations were as follows: IG (n = 86), 1.19; PG (n = 62), 0.58; and GG (n = 38), 0.37. Severity of illness significantly increased over the generations (F = 8.39, 2 df, P = .0003). Pairwise comparisons using the Student-Newman-Keuls test revealed that the significant difference was between the IG and the PG and GG. PG and GG severity means were not significantly different from each other.

Age at First Hospitalization

The survival curve (fig. 2) shows that subjects were first hospitalized for psychosis at progressively younger ages, across generations (χ^2 = 26.76, P = .0001, 2 df). Sex of subject was not a significant covariate to generation, for age at first hospitalization (increment χ^2 = 0.40, P = .53, 1 df). All of the GG and most (68%) of the PG subjects had achieved an age of 55 years or more, by which time they would have been virtually through the age at risk for schizophrenia (Gottesman and Shields 1982).

Imprinting

There were equal rates of maternal and paternal transmission (six and seven cases, respectively) from the PG to the IG. The mean age at first hospitalization for maternal transmission (PG to IG) was 25.00 years (SD 5.94 years), and that for paternal transmission was 27.64 years (SD 10.63 years), a nonsignificant difference (t = .83, P = .41). Grandparental to parental transmission was predominantly maternal (seven of nine cases), with one unknown and only one example of paternal transmission. Sex of transmitting parent was not a sig-

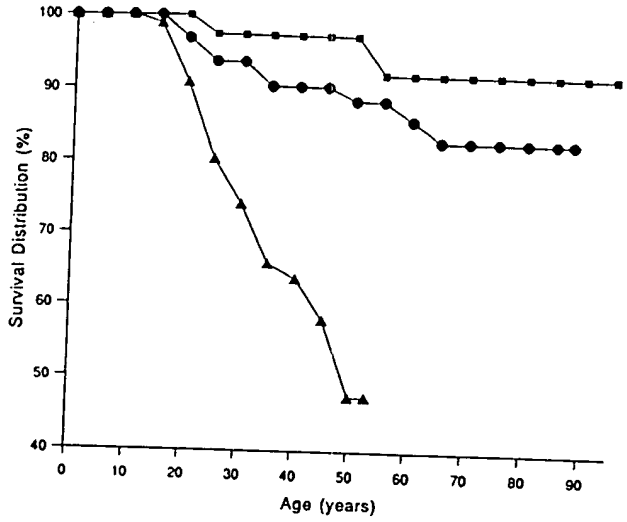


Figure 2 Survival curves for age at first hospitalization for psychotic illness, comparing IG (▲), PG (●), and GG (■). Annual percent surviving without hospitalization is plotted for every 5th year. Observation of individuals ended at their age of death or, if they were living, at their current age.

nificant covariate in the survival analysis examining age at first hospitalization (increment $\chi^2 = 0.32$, $P = .57$, 2 df).

Discussion

The results suggest that familial schizophrenia exhibits anticipation. All of the families studied showed this phenomenon, manifest as increasing rates of hospitalized psychotic illness, worsening severity of illness, and/or earlier age at onset, across successive generations (fig. 1). These findings are consistent with differences in rates of hospitalization and age-at-onset data for parent-child pairs in studies of schizophrenia over the century (Mott 1910; Kay 1963; Penrose 1971; Decina et al. 1991). In each of these studies, rates of hospitalization for psychosis were lower for antecedent generations, and age at onset for parents was significantly later than that for offspring. Investigations of ancestors and extended families also support these findings (Karlsson 1966; Odegaard 1972; Wetterberg and Farmer 1991). As well, less severe psychotic illnesses (e.g., affective disorders) are consistently more common in the generation antecedent to schizophrenic probands (Slater and Cowie 1971; Bleuler 1978). These results complement reported morbid risk of schizophrenia for parents, which is almost half that for siblings (Gottesman and Shields 1982). While alternative reasons, including selection biases such as reduced fertility in earlier-onset schizophrenia, have been proposed to explain these clinical observations, they are all consistent with the phenomenon of anticipation in schizophrenia.

In contrast, findings from the current study are only suggestive of sex-specific differences in the transmission of schizophrenia. However, the possibility of an excess maternal over paternal transmission in schizophrenia is consistent with trends found both recently by others (Sharma et al. 1993) and in studies of large data sets in the older literature (Penrose 1971; Slater and Cowie 1971). Fertility may be especially low in male patients with schizophrenia (Gottesman and Shields 1982), and in the current study this could be the reason for the high rate of maternal transmission from the GG lines to the PG lines. If this were the case, however, one would have expected predominantly maternal inheritance from the PG to the IG lines; but maternal and paternal rates were equal. Comparable transmission patterns have been found to be due to greater variation of trinucleotide repeat length after female meiosis in myotonic dystrophy (Lavedan et al. 1993). The effect may be subtle, requiring larger samples to demonstrate imprinting in schizophrenia.

There are several possible biases that can explain results that indicate anticipation (Penrose 1948). First, subjects could have died before expressing the mutation. However, all of the GG and most of the PG lived beyond age 55 years. Also, significantly different hospitalization rates across generations remained in the current study, even when half of those who had moved or died before age 40 years were assigned affected status. Second, reduced fertility of individuals with earlier onset of schizophrenia could cause preferential ascertainment of parents with later onset. This bias should be minimized in the current study, because (1) few parents were affected with psychosis and (2) siblines were large in all three generations, providing multiple opportunities for detection of affecteds, regardless of their fertility. Third, subjects in the IG could have been too young to yet express a late-onset form of psychosis, which would attenuate the age-at-onset findings. Since one-half of the subjects in the IG were over age 40 years, most were beyond the period of highest risk. Even if more new cases of psychosis subsequently arose, this would most likely occur in the IG and would only serve to strengthen the present study's findings with respect to differential rates of illness.

In contrast to major depression (Gershon et al. 1987), there is no evidence for a cohort effect in schizophrenia. However, secular trends, such as improved detec-

tion, that could, over time, lead to higher hospitalization rates and/or younger age at first hospitalization for psychosis are important to consider, since these could have influenced the principal findings of the current study. In the literature (Mott 1910; Kay 1963; Penrose 1971; Decina et al. 1991), examination of ages at first hospitalization did not reveal secular trends to younger age over the century. In the current study, arbitrary increases in hospitalization incidence assigned to the PG or GG did not change the observation of anticipation. Specific factors that may influence secular trends in hospitalization, including drug abuse and psychosocial stressors, do not appear to have played a role in the sample studied. Only three hospitalized subjects in the IG had a history of stimulant or hallucinogen use, drugs that may in some cases precipitate a psychotic illness. On the basis of direct interviews, it appeared that psychosocial stressors endured by PG and GG, such as the World Wars and Great Depression, were more severe than those faced by the IG.

Another possibility is that the family-history method tends to underestimate rates of psychiatric disturbance, particularly in relatives who are dead or less known (Andreasen et al. 1977). The consequence could be that actual rates for schizotypal traits could be higher than those found, particularly for PG and GG. However, results for severe illness requiring hospitalization would remain unchanged. Another factor that could have compounded errors in the PG and GG was the use of IG siblines connected at the parental or grandparental level. However, both the fact that results remained the same when only one ascending line from each kindred was examined and the fact that at least 13 affected siblines had resulted from eight originating GG lines support the finding of anticipation.

A limitation that is important to consider in studies of common illnesses is the potential for assortative mating to cause an apparently increased prevalence of illness in offspring. This possibility was minimized by selecting unilineal pedigrees with no evidence of schizophrenia or related disorders in the married-in person, their siblings, or parents. Although individuals marrying in could have been nonexpressing carriers of the disorder with nonexpressing close relatives, the likelihood appears small that assortative mating could account for the results in the current study. Because the families studied were selected because of their autosomal dominant-like inheritance and large sibships, the results may not be generalizable to schizophrenia in the general population, although they are consistent with observations from large population-based samples (Penrose 1971; Bleuler 1978). Also, the schizophrenia in the subject families may be a particularly severe form, which could exaggerate the findings. However, the mean age at onset for the IG is similar to others' results (Mott 1911; Gottesman and Shields 1982; Decina et al. 1991; Sharma et al. 1993). Data on specific symptom patterns suggest that the familial schizophrenia in the present families is comparable in nature and in severity to samples drawn from the general population (Bassett et al. 1993).

Despite the possible biases and limitations, the weight of evidence from both the current investigation and the literature is consistent with the finding of anticipation in familial schizophrenia. Other families with schizophrenia should be examined for anticipation and possible accompanying maternal imprinting phenomena, to confirm the current study's findings. Although published pedigrees consistently show evidence of anticipation (Karlsson 1966; Wetterberg and Farmer 1991), complete ascertainment of other large kindreds with contemporary reliable diagnostic assessments would be useful. However, the most exciting possibilities—and the confirmation of the clinical observations of the current study—lie in the search for expanding trinucleotide repeats and other DNA sequence mutations in schizophrenia. New methods becoming available to detect these mutations (Orr et al. 1993; Schalling et al. 1993) will complement and may accelerate the search for pathological genes in linkage studies. In addition, the current study has implications for the genetic model used in linkage studies. Parameters, particularly penetrance, which would vary according to generation, may need to be modified to reflect the effects of anticipation and possibly imprinting. In the light of clinical evidence for anticipation, these strategies represent real promise for deciphering the genetics of schizophrenia.

Acknowledgments

The authors are grateful to the families for their cooperation; to J. McAlduff, R.N., and A. Bury, B.A., for invaluable assistance; and to N. Risch, Ph.D., for comments on the manuscript. This work was supported in part by funding from the Scottish Rite Schizophrenia Research Program, Ian Douglas Bebensee Foundation, Ontario Mental Health Foundation, and Medical Research Council of Canada.

References

Andreasen N, Endicott J, Spitzer R, Winokur G (1977) The family history method using diagnostic criteria: reliability and validity. Arch Gen Psychiatry 34:1228–1235

Bassett AS (1991) Linkage analysis of schizophrenia: challenges and promise. Soc Biol 38:189–196

Bassett AS, Collins EJ, Nuttall S, Honer WG (1993) Positive and negative symptoms in families with schizophrenia. Schizophr Res 11:9–19

Bleuler M (1978) The schizophrenic disorders: long-term patient and family studies. Yale University Press, New Haven

Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, et al (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. Cell 68:799–808

Caskey CT, Pizzuti A, Fu Y, Fenwick RG Jr, Nelson DL (1992) Triplet repeat mutations in human disease. Science 256:784–788

Decina P, Mukherjee S, Lucas L, Linder J, Horwath E (1991) Patterns of illness in parent-child pairs both hospitalized for either schizophrenia or a major mood disorder. Psychiatry Res 39:81–87

Flint J (1992) Implications of genomic imprinting for psychiatric genetics. Psychol Med 22:5–10

Fu Y-H, Pizzuti A, Fenwick RG Jr, King J, Rajnarayan S, Dunne PW, Dubel J, et al (1992) An unstable triplet repeat in a gene related to myotonic muscular dystrophy. Science 255:1256–1258

Gershon ES, Hamovit JH, Guroff JJ, Nurnberger JI (1987) Birth-cohort changes in manic and depressive disorders in relatives of bipolar and schizoaffective patients. Arch Gen Psychiatry 44:314–319

Gottesman II, Shields J (1982) Schizophrenia: the epigenetic puzzle. Cambridge University Press, Cambridge

Harper PS, Harley HG, Reardon W, Shaw DJ (1992) Anticipation in myotonic dystrophy: new light on an old problem. Am J Hum Genet 51:10–16

Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72:971–983

Karlsson JL (1966) The biologic basis of schizophrenia. Charles C Thomas, Springfield, IL

Kay DWK (1963) Late paraphrenia and its bearing on the aetiology of schizophrenia. Acta Psychiatr Scand 39:159–169

Kendler KS, Gruenberg AM, Tsuang MT (1985) Psychiatric illness in first-degree relatives of schizophrenic and surgical control patients. Arch Gen Psychiatry 42:770–779

Lavedan C, Hofmann-Radvanyi H, Shelbourne P, Rabes J-P, Duros C, Savoy D, Dehaupas I, et al (1993) Myotonic dystrophy: size- and sex-dependent dynamics of CTG meiotic instability, and somatic mosaicism. Am J Hum Genet 52:875–883

Lowing PA, Mirsky AF, Pereira R (1983) The inheritance of schizophrenia spectrum disorders: a reanalysis of the Danish adoption study data. Am J Psychiatry 140:1167–1171

McInnis MG, McMahon FJ, Chase GA, Simpson SG, Ross CA, DePaulo JR Jr (1991) Anticipation in bipolar affective disorder. Am J Hum Genet 53:385–390

Mott FW (1910) Hereditary aspects of nervous and mental diseases. Br Med J 2:1013–1020

——— (1911) Heredity and insanity. Lancet 3:1251–1259

Odegaard O (1972) The multifactorial theory of inheritance in predisposition to schizophrenia. In: Kaplan AR (ed) Genetic factors in schizophrenia. Charles C Thomas, Springfield, IL, pp 256–275

Orr HT, Chung M, Banfi S, Kwiatkowski TJ Jr, Servadio A, Beaudet AL, McCall AE, et al (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nature Genet 4:221–226

Penrose LS (1948) The problem of anticipation in pedigrees of dystrophia myotonica. Ann Eugenics 14:125–132

——— (1971) Critical survey of schizophrenia genetics. In: Howells JG (ed) Modern perspectives in world psychiatry. Oliver & Boyd, Edinburgh, pp 3–19

Risch N (1990) Linkage strategies for genetically complex traits. I. Multilocus traits. Am J Hum Genet 46:222–228

Schalling M, Hudson TJ, Buetow KH, Housman DE (1993) Direct detection of novel expanded trinucleotide repeats in the human genome. Nature Genet 4:135–139

Sharma T, Brynjolfsson J, Butler R, Read T, Murphy P, Sigmundsson T, Petursson H, et al (1993) Genomic imprinting in the Icelandic/British schizophrenia study: preliminary results. Schizophr Res 9:124–125

Slater E, Cowie V (1971) The genetics of mental disorders. Oxford University Press, London

Sutherland GR, Richards RI (1992) Anticipation legitimized: unstable DNA to the rescue. Am J Hum Genet 51:7–9

Verkerk AJMH, Pieretti M, Sutcliffe JS, Fu Y-H, Kuhl DPA, Pizzuti A, Reiner O, et al (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 65:905–914

Wetterberg L, Farmer AE (1991) Clinical polydiagnostic studies in a large Swedish pedigree with schizophrenia. Eur Arch Psychiatry Clin Neurosci 240:188–190

The Journal of Cell Biology

Journal of Experimental Medicine Online

J. Cell Biol.

0021-9525/97/09/1089/15 $2.00
Volume 138, Number 5, September 8, 1997 1089-1103

Abstract of this Article
Reprint (PDF) Version of this Article
Similar articles found in:
JCB Online
PubMed
PubMed Citation
This Article has been cited by:
other online articles
Search Medline for articles by:
Cimato, T. R. || Aletta, J. M.
Alert me when:
new articles cite this article
Download to Citation Manager

Nerve Growth Factor-specific Regulation of Protein Methylation during Neuronal Differentiation of PC12 Cells

Thomas R. Cimato,* Murray J. Ettinger,* Xianbo Zhou,* and John M. Aletta‡

* Department of Biochemistry, ‡ Department of Pharmacology and Toxicology, University at Buffalo School of Medicine and Biomedical Sciences, State University of New York, Buffalo, New York 14214

Abstract

Protein methylation is a posttranslational modification that can potentially regulate signal transduction pathways in a similar manner as protein phosphorylation. The role of protein methylation in NGF signaling was examined by metabolic labeling of PC12 cell proteins with L-[methyl-^{3}H]methionine and by in vitro labeling of cell proteins with L-[methyl-^{3}H]S-adenosylmethionine. Effects of NGF were detected within 15 min. Methyl-labeled proteins were resolved by one and two dimensional SDS-PAGE. NGF affected the methylation of several 68-60-kD proteins (pI 5.8-6.4) and 50-kD proteins (isoelectric point pH 6.7-6.8 and 5.8-6.2). Several NGF-induced changes in methylation increased over several hours and through 4 d. Moreover, methyl labeling of several specific proteins was only detected after NGF treatment, but not in nontreated controls. The effects of NGF on protein methylation were NGF specific since they were not observed with EGF or insulin. A requirement for protein methylation for neurite outgrowth was substantiated with either of two methylation inhibitors: dihydroxycyclopentenyl adenine (DHCA) and homocysteine. DHCA, the more potent of the two, markedly inhibits protein methylation and neurite

EXHIBIT B

outgrowth without affecting cell growth, NGF-induced survival, cell flattening, or several protein phosphorylations that are associated with early signaling events. Removal of DHCA leads to rapid protein methylation of several proteins and concurrent neurite outgrowth. The results indicate that NGF regulates the methylation of several specific proteins and that protein methylation is involved in neurite outgrowth from PC12 cells.

PROTEIN methylation is a posttranslational modification that may be used to regulate signal transduction and differentiation pathways by mechanisms that are analogous to regulation by protein phosphorylation (Hrycyna and Clarke, 1993; Rando, 1996). Although a specific role for protein methylation in prokaryote chemotaxis is well established (Shapiro et al., 1995), possible roles for protein methylation in eukaryotic signaling mechanisms have not been extensively explored. Protein carboxyl methylations are reversible, and regulatory roles for carboxyl methylation have been proposed for chemoattractant responses in neutrophils (Philips et al., 1993, 1995), insulin secretion from pancreatic islets (Metz et al., 1993), and photoreceptor signal transduction (Parish et al., 1995). Among signaling proteins that are known to be carboxyl methylated are the Ras and Rho family of small G-proteins (Hrycyna and Clarke, 1993; Rando, 1996), γ subunits of heterotrimeric G-proteins (Philips et al., 1993; Rando, 1996), and the catalytic subunit of protein phosphatase 2A (Lee and Stock, 1993; Favre et al., 1994; Xie and Clarke, 1994). Protein phosphatase 2A is also demethylated by a specific protein carboxyl methylesterase (Lee et al., 1996). Many proteins are known to be N-methylated at arginine, lysine, or histidine residues including cytoskeletal proteins (actin and myosin), nuclear proteins (nucleolin, fibrillarin, histones, heterogeneous nuclear RNPs), the multifunctional calcium binding protein, calmodulin, and FGF-2. The physiological functions of methylation, however, remain largely unknown. Moreover, potential mechanisms for regulating protein methylation within growth factor signaling pathways remain to be explored.

Methylation pathways use S-adenosylmethionine (SAM)[1] as the universal methyl donor for methyltransferase-catalyzed methylation of proteins and other methyl acceptors. The role of methylation in NGF signal transduction was previously studied using high concentrations (millimolar) of inhibitors of methyltransferases that use SAM (Seeley et al., 1984; Kujubu et al., 1993). These attempts to examine the role of methylation in cell signaling were somewhat compromised by the lack of specific, nontoxic inhibitors of methylation. In addition to the methylated acceptor, S-adenosylhomocysteine (SAHcy) is a product of all methyltransferase reactions. SAHcy is a strong competitive inhibitor of SAM (Hildeshein et al., 1972) and is normally removed by hydrolysis in a reversible reaction (De la Haba and Cantoni, 1959) catalyzed by S-adenosylhomocysteine hydrolase (SAHH). Thus, inhibition of SAHH offers an alternative means to inhibit methyltransferases. This approach was used in the present work to study protein methylation during NGF signaling. The crucial role of SAHH and methylation in development is evident from the nonagouti (a^x) mutation. The lethality of this mutation is due to a deletion of the SAHH gene, and embryonic development is arrested in the preimplantation blastula stage (Miller et al., 1994).

The PC12 clonal cell line was used to examine the role of protein methylation in NGF signal transduction. PC12 cells are derived from a rat pheochromocytoma and serve as a model of neuronal differentiation. When exposed to NGF, PC12 cells assume many of the features of sympathetic neurons including cell

cycle arrest, survival in serum-free medium, and elaboration of long neurites (Greene and Tischler, 1976; Tischler and Greene, 1978). Moreover, PC12 cells also respond to other growth factors. For example, in the presence EGF, PC12 cells do not differentiate into neuronlike cells, but do increase their mitotic rate (Huff and Guroff, 1981). Thus, PC12 cells are useful to study specific growth factor signaling mechanisms. The specific signaling events and proteins involved in neurite outgrowth have not been fully elucidated. For example, NGF and EGF both activate several signaling proteins including phospholipase C, phosphoinositide (PI)3-kinase, extracellular signal-related kinase (ERKs) (MAP kinases), and Ras (Kaplan and Stephens, 1994; Marshall, 1995). However, NGF, but not EGF, stimulates neurite outgrowth in PC12 cells. One hypothesis for the specificity of NGF actions involves kinetic differences in the activation of both Ras and ERK (Marshall, 1995). NGF stimulation of PC12 cells results in prolonged activation of both Ras and ERK activity, while EGF causes a transient rise and fall in Ras and ERK activities (Qui and Green, 1992). Specificity may also arise from NGF-specific signaling pathways that are not activated by EGF and other mitogens (Chao, 1992; Kaplan and Stephens, 1994; Peng et al., 1995). The results reported here indicate that signaling pathways involving protein methylation may contribute to NGF specificity.

Previous studies by Seeley et al. (1984) have implied that methylation is necessary for neurite outgrowth in PC12 cells. Kujubu et al. (1993) and Haklai et al. (1993) reported that methylation of small G-proteins is regulated by NGF in PC12 cells. Najbauer and Aswad (1990) have also characterized several methyl-arginine containing proteins in PC12 cells. The effects of NGF on protein methylation and the effect of inhibition of protein methylation on neurite outgrowth are, however, incompletely characterized and have been difficult to interpret due to the lack of specificity of the methylation inhibitors used in previous studies. The present work extends previous findings by examining the methylation of specific cellular proteins and neurite outgrowth before and after inhibition of the protein methylation pathway. 9-(*trans*-2′, *trans*-3′-dihydroxycyclopent-4-enyl)-adenine (DHCA), a specific, mechanism-based inhibitor of SAHH, inhibits protein methylation and greatly decreases neurite outgrowth. The results indicate that NGF, but not EGF or insulin, regulates the methylation of several specific proteins and that protein methylation is required for neurite outgrowth from PC12 cells.

Materials and Methods

Reagents

NGF was purified from male mouse submaxillary glands as described previously (Mobley et al., 1976). Insulin, L-homocysteine thiolactone, anisomycin, leupeptin, PMSF, aprotinin, and monoclonal phosphotyrosine antibody (clone PT-66) were purchased from Sigma Chemical Co. (St. Louis, MO). Homocysteine was prepared by incubating L-homocysteine thiolactone with 40 mM NaOH at 37°C for 30 min. EGF was obtained from Upstate Biotechnology Inc. (Lake Placid, NY). DHCA (kindly provided by R.T. Borchardt, University of Kansas, Lawrence, KS) was prepared as 1 and 100 mM stock solutions in DMSO. Erythro-9-(2-hydroxy-3-nonyl) adenine was kindly provided by D. Porter (Burroughs Wellcome, Research Triangle Park, NC). L-[methyl-^{3}H]methionine (71.4 Ci/mmol), L-[methyl- ^{3}H]SAM (70 Ci/mmol), [^{14}C]adenosine (59.8 mCi/mmol), and carrier-free [^{32}P]orthophosphate were purchased

from Dupont-NEN (Boston, MA). ^{125}I-donkey anti-rabbit IgG was purchased from Amersham Corp. (Arlington Heights, IL). Protein A-Sepharose (PAS) and ampholines were obtained from Pharmacia Fine Chemicals (Piscataway, NJ). Donor horse serum and fetal bovine serum were from JRH Biosciences (Lenaxa, KS). Cell culture medium, penicillin, and streptomycin were obtained from Gibco Laboratories (Grand Island, NY). Restriction endonuclease MspI was from New England Biolabs Inc. (Beverly, MA), and nuclease P1 from United States Biochemical Corp. (Cleveland, OH).

Cell Culture and Bioassays

Stock cultures of PC12 cells were grown on collagen-coated tissue culture dishes (Falcon Plastics, Cockeysville, MD) in RPMI-1640 medium supplemented with 10% heat-inactivated horse serum and 5% fetal bovine serum at 35°C and 7.5% CO_2, as previously described (Greene et al., 1987). For neurite outgrowth studies, PC12 cells (3-5 × 10^5 cells) were plated onto collagen-coated 35-mm plastic tissue culture dishes and cultured in RPMI-1640 medium containing 1% heat-inactivated horse serum and 50 ng/ml NGF. Neurites >20 μm were counted as neurite-bearing cells. At least 100 cells from each experimental condition were scored from randomly chosen fields.

For the study of proliferation and survival in the presence and absence of DHCA, naive PC12 cells were plated on collagen-coated, 24-well tissue culture dishes at a density of 10^5 cells per well. 1 d after plating, cells were washed three times in RPMI-1640 medium to remove serum and then placed either in complete medium (RPMI plus 10% horse serum and 5% fetal bovine serum), in RPMI plus 50 ng/ml NGF, or in RPMI alone. Where indicated, 1 μM DHCA was included in the culture medium. The number of viable cells was determined by counting intact nuclei using a hemacytometer (Soto and Sonnenschein, 1985) immediately after washing the cultures with RPMI at time intervals of 1, 3, and 5 d later.

For the study of neurite regeneration, PC12 cells were treated with NGF for 7-14 d. NGF was then thoroughly washed away from the cells, and the cells were detached from the culture dish by trituration. Cells were then replated in the presence or absence of NGF, and neurite outgrowth was scored 24 h later.

Metabolic Radiolabeling of Proteins

To assess endogenous protein methylation under various experimental conditions, PC12 cells were cultured in 1% horse serum for at least 16 h before metabolic radiolabeling for 6 h with 100 μCi/ml L-[methyl-^{3}H]methionine in the presence of 10 μM anisomycin at 35°C in a CO_2 incubator. The methionine/cysteine content of RPMI-1640 was reduced by 80% during the labeling period. Growth factors (NGF, EGF, or insulin) and/or 1 μM DHCA were added for the indicated times. At the conclusion of the labeling period, cells were washed three times with 1 ml of PBS (35°C), harvested in 200 μl SDS-PAGE lysis buffer, and held in a boiling water bath for 5 min. Total radiolabeled protein in each lysate was determined by TCA precipitation and liquid scintillation spectrometry. The averages of the determinations are given under Results as the means ± SEM. Aliquots of each cell lysate containing equal TCA-precipitable counts per minute were subjected to SDS-PAGE as described below. For analysis of protein methylation by two-dimensional (2D) IEF × SDS-PAGE, the cells were harvested in IEF lysis buffer (Aletta and Greene, 1987) without heating the sample.

To assess changes in protein phosphorylation, PC12 cell proteins were metabolically radiolabeled as previously described (Aletta, 1996) in a Hepes-buffered Krebs-Ringer solution containing 0.1% D-glucose. When the effect of DHCA was examined, the drug was added to cultures 30 min before addition of the radioisotope. The labeling was carried out with 50 μCi/ml [^{32}P]orthophosphate at 35°C in room air for 2 h. At the end of the labeling period, cells were washed three times with 1 ml of PBS (35°C), harvested in SDS-PAGE lysis buffer, and held in a boiling water bath for 5 min. Total radiolabeled protein in each lysate was determined by TCA precipitation and liquid scintillation spectometry.

In Vitro Protein Methylation

To measure protein methylation in subcellular fractions, cells were washed three times in PBS (4°C) and harvested by scraping with a rubber policeman in ice-cold homogenization buffer (100 mM Tris, pH 8.0, 1 mM EDTA, 2 μM PMSF, and 10 μg/ml leupeptin). After homogenization at 4°C in a Dounce homogenizer, the homogenates were centrifuged at 4°C for 10 min at 500 *g* to obtain a crude nuclear fraction. The supernatant was centrifuged at 20,000 *g* for 5 min at 4°C. The resulting supernatant was centrifuged at 100,000 *g* for 90 min at 4°C to obtain the cytoplasmic fraction. The pellet of the 100,000 *g* spin was washed three times in homogenization buffer and resuspended in 150 μl of homogenization buffer to obtain the membrane fraction. Protein in each fraction was determined spectrophotometrically (Bradford, 1976). Equal amounts of protein (125 μg) from each fraction were incubated with 4.25 μCi L-[methyl-^{3}H]SAM in a total volume of 50 μl for 1 h at 37°C. The labeling reaction was stopped by adding 12.5 μl of 5× SDS-PAGE sample buffer (0.3 M Tris-HCl, pH 6.8, 45% glycerol, 1.4 M 2-mercaptoethanol, 10% SDS, 0.001% bromophenol blue) and holding the samples in a boiling water bath for 5 min. In vitro-labeled proteins from the cell fractions were then analyzed by 7.5-15% gradient SDS-PAGE as described below.

Gel Electrophoresis

Discontinuous SDS-PAGE (Laemmli, 1970) was performed with 19-cm separating gels composed of polyacrylamide gradients of 6-12, 7.5-15, or 8.5-15%, depending upon the experiment. Gels were fixed, stained with Coomassie blue, and then destained. Gels containing proteins labeled with [^{32}P]orthophosphate were dried and placed in contact with Kodak XAR film to produce an autoradiographic image. Gels containing tritium- labeled proteins were prepared for fluorography by washing the gel for 1 h in three changes of deionized water, followed by immersion in 1 M sodium salicylate for 30 min (Chamberlain, 1979). After drying, the gels were exposed to preflashed Kodak XAR film (Laskey and Mills, 1975) and stored at −70°C (Bonner and Laskey, 1974) with an intensifying screen. Quantitative comparisons of methyl-labeled proteins were obtained by scanning fluorograms into an analysis program (Molecular Analyst; Bio Rad Laboratories, Hercules, CA).

For 2D IEF × SDS-PAGE, PC12 cell proteins were labeled as described above and harvested in a lysis buffer appropriate for IEF (Aletta and Greene, 1987). Equal TCA-precipitable counts per minute of cell lysates were subjected to IEF with pH 5-7, and 3.5-10 ampholines at a ratio of 4:1, respectively. Proteins in IEF gels were further resolved by SDS-PAGE (the second-dimension) using 12-cm separating gels composed of 7.5-15% polyacrylamide gradient. A standard IEF gel containing only lysis solution was used to determine the pH range. Fluorographic images of tritium-labeled proteins were generated as described

above.

DNA Methylation

Total methylation of cytosines in DNA was determined by first digesting PC12 cell DNA with MspI, which cleaves methylated or unmethylated CCGG sequences (Bestor et al., 1984). The products are 5′ labeled with [^{32}P]ATP, and digested with nuclease P1. Methyl-cytosines are resolved from unmethylated cytosines by chromatography in isobutyric acid/water/ ammonium hydroxide (66:33:1) and detected by autoradiography (Bestor et al., 1984). Liquid scintillation spectrometry was used to quantify the results.

Immunoprecipitation

Tyrosine phosphorylation of Trk in the presence or absence of DHCA was assessed by immunoprecipitation followed by Western blotting. Cells were treated with 100 ng/ml NGF for 5 min, with or without 1 μM DHCA. Cells were then washed three times in ice-cold PBS and harvested in 1 ml of a lysis buffer containing 1% Triton X-100, 150 mM NaCl, 50 mM Tris-HCl, pH 8.0, 25 mM NaF, 5 mM EGTA, 5 mM EDTA, 2 μM PMSF, and 100 U/ml aprotinin. Insoluble material was removed by centrifugation at 4°C for 10 min at 13,000 *g*. Samples were precleared with 6 mg PAS for 2 h followed by centrifugation for 10 min at 13,000 *g*. Anti-phosphotyrosine monoclonal antibody (clone PT-66) was added to equal amounts of lysate protein for 2 h on a rotating platform at 4°C followed by addition of 3 mg of PAS and incubation on the rotating platform at 4°C for 1 h. PAS beads were then recovered by centrifugation at 13,000 *g* for 10 min, and washed three times with 1 ml of 1% Triton X-100 lysis buffer followed by two 1-ml washes in lysis buffer without Triton X-100. The PAS beads were then resuspended in SDS-PAGE sample buffer and held in a boiling water bath for 5 min. The precipitated material was resolved by SDS-PAGE (7.5% acrylamide) followed by transfer to Immobilon P membrane (Millipore Corp., Milford, MA). The blot was probed with 1 μg of rabbit polyclonal Trk antiserum (Santa Cruz Biotechnology, Santa Cruz, CA) for 2 h at room temperature followed by incubation with ^{125}I-donkey anti-rabbit IgG. The dried blot was then exposed to Kodak XAR film at −70°C with an intensifying screen.

SAHH Activity

PC12 cells were plated on 150-mm dishes (2×10^7 cells) in the presence of 100 ng/ml NGF plus 10 nM to 3 μM DHCA, or without added DHCA for 7 d. Cells were then washed three times in ice-cold PBS and harvested by scraping in a potassium phosphate buffer, pH 7.0 (25 mM KH_2PO_4, 25 mM K_2HPO_4, 1 mM dithiothreitol, 0.5 μM PMSF, and 10 μg/ml leupeptin). Cells were homogenized in a Dounce homogenizer, and the cell nuclei and debris were removed by centrifugation at 13,000 *g*. Cytosol was then prepared by centrifugation at 100,000 *g* for 90 min and the protein concentration determined. SAHH activity was determined in the synthesis direction by a TLC method described previously by Hershfield (1979).

Results

NGF-specific Induction of Changes in the Pattern of PC12 Cell Methylated Proteins

Previous studies have implicated regulation of protein methylation in the early events of NGF-mediated signal transduction (Seeley et al., 1984; Kujubu et al., 1993). To examine this possibility in greater detail, protein methylation patterns were assessed after metabolic radiolabeling of cellular proteins with L-[methyl-^{3}H]methionine. Fluorograms of the labeled proteins generated from SDS-PAGE and 2D IEF × SDS-PAGE were used to detect specific changes in protein methylation after NGF treatment. The cellular pool of SAM, the predominant methyl donor in all cells, was radiolabeled using L-[methyl-^{3}H]methionine in control PC12 cells, and PC12 cells treated with either 50 ng/ml NGF, 5 nM EGF or 1 μM insulin. These concentrations and growth factors were chosen based on the biological effects produced by each in PC12 cells. NGF at 50 ng/ ml produces the maximum neurite outgrowth response. EGF at 5 nM enhances cell proliferation (Huff et al., 1981) and 1 μM insulin is commonly used to effect cell survival in serum-free media (Rukenstein et al., 1991). We have verified that receptors for all three of these growth factors are present on the PC12 cells used in these studies, by observation of the aforementioned biological effects at the doses indicated (data not shown). Radiolabeled methionine equilibrates with the cellular SAM pool within 20 min (Chelsky et al., 1985). A protein synthesis inhibitor, anisomycin (10 μM), was present during the metabolic radiolabeling to prevent incorporation of L-[methyl-^{3}H]methionine into newly synthesized proteins.

In each of five independent trials, all three trophic factors produced a net increase in methyl group incorporation into PC12 cell protein. After 6 h of treatment, EGF and insulin lead to ~50% greater incorporation (±0.2, SEM) and NGF to a 31% increase (±0.1, SEM). Despite the slightly larger effects of EGF- and insulin-promoted increases in total methyl-^{3}H-labeled protein, NGF treatment consistently yields protein methylation patterns exhibiting more pronounced changes in specific methylated proteins (Fig. 1). Several, indistinctly resolved proteins that migrated between 24 and 20 kD showed increased methyl labeling relative to total protein methylation after 6 h NGF treatment (Fig. 1, *top right*). Among the methylated species detected in control cell lysates, were nine 68-60-kD proteins with isoelectric point (pI) values of 5.8-6.4 (Fig. 1, *A-G*, *, and +). After treatment with 50 ng/ml NGF for 6 h, the proteins labeled *A*, *, and + in Fig. 1 showed little or no reproducible changes in methylation relative to controls. Although there was a net increase in total protein methylation, both increases and decreases in the methylations of specific proteins relative to total protein methylation were detected in response to NGF. Decreased labeling of proteins *B*, *D*, and *E* was detected with NGF (Fig. 1 *b*). Methyl labeling of proteins *C*, *F*, and *G* increased markedly, and methyl group incorporation into two proteins (N and N′) was detected only after 6 h in NGF-treated cells, but not in controls (Fig. 1 *b*).

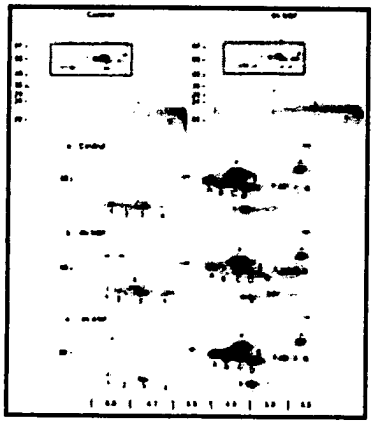


Fig. 1. NGF induces specific changes in the pattern of in vivo protein methylation observed by 2D IEF × SDS-PAGE. PC12 cell proteins were metabolically labeled with L-[methyl-^{3}H]methionine (100 μCi/ml) in the presence of anisomycin (10 μM) for 6 h. Whole cell lysates containing equal TCA-precipitable cpm (800,000) were separated by IEF. IEF gels were loaded onto 8.5- 15% gradient gels for the separation of methylated proteins by molecular mass. Fluorographic images generated from the control and 6 h NGF (50 ng/ml) conditions are illustrated in the top two panels. Molecular mass standards (in kD) are indicated at the left of each panel. The boxed area in the top panels is enlarged below for control, 6 h NGF and 6 h

EGF (5 nM) samples. Fluorograms were generated by exposing x-ray film to the dried gels with an enhancing screen for 10 d at −70°C. The pH gradient of the isoelectric focusing tubes is shown below the enlarged images (*a-c*). Labels (*A*, *, and +) indicate unaffected methylations, while the remaining letters and numbers indicate protein methylations affected by NGF. These results are representative of three independent experiments.
[View Larger Version of this Image (71K GIF file)]

Significant NGF-induced effects on protein methylation were also detected at pI 6.7-6.8 and 5.8-6.2 in the 50-kD region (Fig. 1, *a* and *b*). Labeling of protein 3 was unaffected by NGF. Decreased methyl labeling of protein 1 was detected after 6 h NGF treatment, while proteins 2, 4, and 7 showed increased methyl labeling, and the [methyl- ^{3}H] incorporation into protein 5 was only detected in response to NGF. The effects of NGF on protein methylation are unlikely to be due to incorporation of L-[methyl-^{3}H]methionine into newly synthesized proteins because protein synthesis was inhibited by 97 ± 0.1% (n = 3) under the conditions of these experiments. In addition, no significant increases or decreases in the amounts of the specific proteins analyzed in Fig. 1 were detected by 2D IEF × SDS-PAGE of [^{35}S]methionine-labeled proteins in whole cell lysates, in the absence of protein synthesis inhibition (data not shown).

The NGF-induced changes in methyl labeling of specific proteins were not observed with the two other trophic factors tested. After EGF treatment for 6 h, the methylation pattern of the 68-60-kD proteins was quite different from the NGF-induced changes relative to the control, nonstimulated cells (Fig. 1). The pattern of protein methylations with EGF was more similar to the pattern with control cells, but small increases in the methylation of proteins in this relative molecular weight range were reproducibly detected (B, C, D, and E). NGF, on the other hand, decreased the methylation of proteins B, D, and E. The changes in protein methylations were less marked with EGF than NGF and no methylated protein was detected with EGF that was not detected with controls. Moreover, no significant effects of EGF were detected in the 50-kD set of methylated proteins. Insulin had no significant effects on the methylation of the 68-60- or 50-kD proteins (data not shown).

The effects of NGF on the pattern of protein methylation were also examined by in vitro labeling of proteins in subcellular fractions from NGF-treated cells as an independent approach for detecting the protein methylations that are affected by NGF. In vitro labeling provides a complementary method of analysis that does not require a protein synthesis inhibitor. It also indicates the possible cellular locations of protein methylation. PC12 cells were incubated without or with NGF for 15 min, 6 or 16 h. Nuclei (500 *g*), membranes (100,000 *g*, pellet), or cytosol (100,000 *g*, supernatant) were isolated and incubated with [methyl-^{3}H]SAM for 1 h at 37°C, in vitro. Methylated proteins were resolved by 7.5-15% gradient gel SDS-PAGE and detected by fluorography. Changes in protein methylation were detected in nuclei, membranes, and cytosol (Fig. 2). Increases in protein methylation were detected after 15 min of NGF treatment (data not shown) and continued to increase from 6 to 16 h (Fig. 2 *A*). Moreover, several of the proteins that showed increases in methylation have similar molecular weights as those detected by 2D IEF × SDS-PAGE (Fig. 1) after metabolic radiolabeling of intact cells (e.g., 68-64 kD, and 50 kD). Thus, changes in the methylation of proteins in response to NGF was confirmed by two independent approaches. In the nuclear fraction, time-dependent increases (more than twofold) in protein methylation in response to NGF treatment were detected in proteins migrating at 97, 94, 67, and 64 kD. In the membrane fraction, the

methylation of proteins migrating at 50 and 34 kD decreased after 6 or 16 h of NGF. Several proteins in the cytosolic fraction showed increased methylation. Most prominent among these were 114, 94, 50, and 35 kD. Additional NGF-induced effects were observed by the in vitro method. This is most likely due to the higher specific activity of the [methyl-^{3}H]SAM pool in vitro than in the intact cell experiments and the enrichment of proteins by subcellular fractionation. In addition, the in vitro labeling experiments indicate that both the requisite methyltransferase and protein substrate were present at the time of analysis in each of the subcellular fractions that were isolated from NGF-activated cells. Moreover, irrespective of the specific mechanisms responsible for activation of the methylation of specific proteins by NGF, the state of activation was stable during preparation of the fractions.

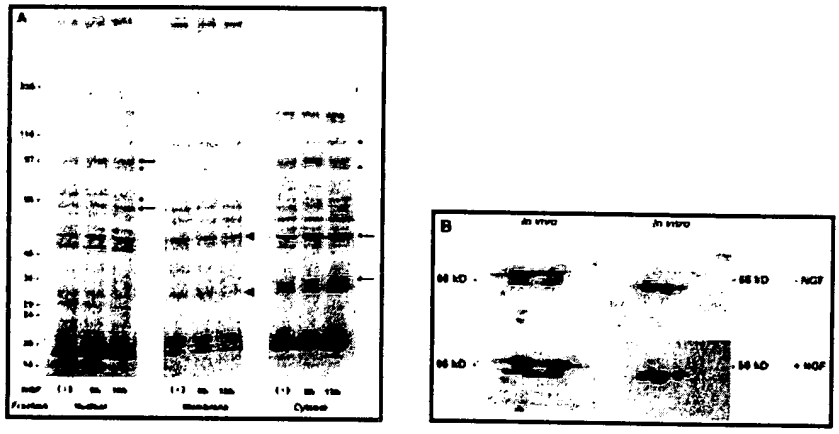


Fig. 2. NGF-treatment of intact cells stimulates in vitro protein methylation in cell-free extracts. (*A*) PC12 cells were incubated in the presence or absence of NGF (50 ng/ml) for 6 or 16 h. Proteins in the nuclear (500 *g*, pellet), membrane (100,000 *g*, pellet), or cytosolic (100,000 *g*, supernatant) fractions were radiolabeled with L-[methyl-^{3}H]SAM for 1 h at 37°C. The labeling reaction was quenched by the addition of 5× SDS-PAGE sample buffer followed by holding the reaction tube in a boiling water bath for 5 min. Labeled proteins were separated by 7.5-15% gradient SDS-PAGE. The image shown is a fluorogram of the dried gel. Asterisks indicate methylated proteins not detectable in extracts from non-NGF-treated control cells. Arrows identify increases of 67% or greater. The arrowheads (membrane fraction) point out decreases of >30%. These results were reproduced in an independent experiment. Specific changes at 16 h were as follows: nuclear proteins at 97 and 64 kD increase 2.4-fold; cytosolic proteins at 50 kD and 35 kD increase 67 and 79%, respectively. (*B*) Cell proteins were metabolically radiolabeled followed by preparation of cytosol (*in vivo*) or the cytosol was prepared first, followed by incubation with [methyl-^{3}H]SAM (*in vitro*). A portion of the 2D IEF × SDS-PAGE fluorograms from control (*−NGF*) versus 6 h NGF treatment (*+NGF*) are displayed. There are five individual protein spots migrating at 64 kD, which are most easily discerned in the in vivo +NGF condition. The 64-kD series of spots in each of the other fluorograms (four spots each) are superimposable with those of the in vivo +NGF condition.
[View Larger Versions of these Images (103 + 40K GIF file)]

To validate this approach further, cytosol was prepared from cells treated with or without NGF after metabolic radiolabeling, and replicate cultures without metabolic radiolabeling were also processed to obtain cytosol for in vitro protein methylation. Comparisons of the 2D IEF × SDS-PAGE fluorograms from each type of preparation indicate that several of the same proteins resolved in the 64 kD range (pI 5.8-6.4)

are similarly increased after NGF treatment by either method of analysis (Fig. 2 *B*). The in vitro data obtained after NGF treatment for 6 h confirm that changes in protein methylation triggered by NGF can occur whether or not protein synthesis has been inhibited. Thus, the in vivo and in vitro experiments independently indicate that NGF produces diverse, marked effects on the regulation of the methylation of several specific proteins.

NGF Affects Protein Methylation during Early and Delayed Signaling

The time dependence of NGF-induced changes in protein methylation was examined to determine the onsets and durations of changes in protein methylation. PC12 cells were incubated with NGF for 15 min to 24 h to examine protein methylation during early stages of NGF signaling and for 4 d to examine changes occurring concurrently with the appearance of neurites. Total protein methylation was assessed by the incorporation of [^{3}H]methyl groups into TCA-precipitable protein during NGF treatment for 24 h (Fig. 3). Significant increases in the total amounts of methyl group incorporation were detected within 15 min after the addition of NGF. The response over the early time course is biphasic with an initial rapid burst of methylation that decays between 1 and 4 h of NGF treatment followed by a second, persistent phase of elevated methylation from 6 to 24 h. The pronounced decrease in methyl group incorporation between 1 and 4 h indicates that the effect of NGF on total protein methylation is likely to be complex and dependent upon multiple factors, including methionine uptake, intracellular compartmentalization, and use of the labeled methyl group in other metabolic pathways. Thus, all comparisons of protein methylation patterns examined by gel electrophoresis were standardized relative to total methyl-^{3}H-labeled protein.

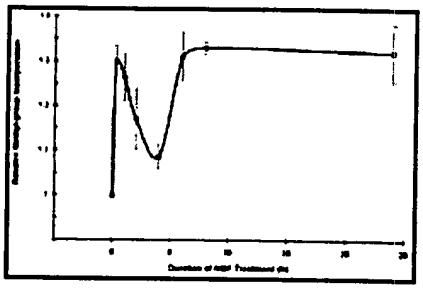


Fig. 3. The time dependence of total methyl group incorporation into proteins in response to NGF. PC12 cell proteins were metabolically labeled with L-[methyl-^{3}H]methionine (100 μCi/ ml) for 6 h in the presence of anisomycin (10 μM). NGF (50 ng/ ml) was added to the cell cultures for the times indicated. Cells were lysed in SDS-PAGE sample buffer. Total methyl labeling of proteins (counts per minute) was measured for each condition after TCA precipitation of an aliquot of the sample. The relative amount of the ^{3}H-labeled methyl groups in lysate proteins at each time point is given with respect to nontreated control cells. The data shown are the means ± SEM for five independent experiments.
[View Larger Version of this Image (10K GIF file)]

Although the resolution of changes in protein methylation in whole cell lysates is reduced on one dimensional SDS-PAGE gels, this method was used to observe the time course of changes in the set of proteins migrating at 64-62 kD. Gradient SDS-PAGE and fluorographic analysis of methyl-^{3}H-labeled proteins in whole cell lysates indicated that there are increases in the methylation of the proteins in this region of the gel, detectable within 1 h after the addition of NGF and persisting for at least 8 h (Fig. 4 *A*). In addition, when using an 8.5-15% acrylamide gradient, a reproducible NGF-specific change in a 34-kD protein is evident at 3 and 6 h of NGF treatment (Fig. 4 *B*), but not at times earlier than 2 h (data not shown).

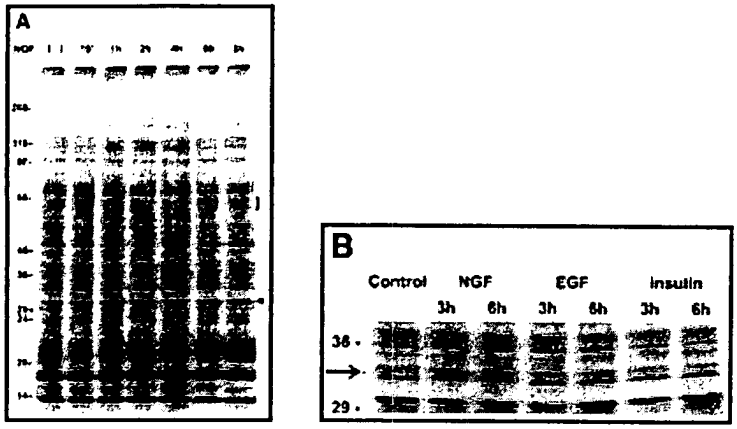


Fig. 4. Changes in the pattern of protein methylation are NGF specific and time dependent. (*A*) PC12 cell proteins were metabolically labeled with L-[methyl-^{3}H]methionine (100 μCi/ml) for 6 h in the presence of anisomycin (10 μM) with or without NGF (50 ng/ml) for the times indicated. Whole cell lysates containing equal TCA-precipitable cpm (300,000) were loaded in each lane. Proteins were separated by 7.5-15% gradient SDS-PAGE. The image shown is a fluorogram of the dried gel after exposure to x-ray film for 3 d at −70°C with an intensifying screen. The bracket indicates the migration position of a 64-62-kD complex of proteins that increase in a time-dependent manner. The arrowhead indicates the migration position of an invariant protein band, the density of which does not change by more than a few percent over the time course. By this measure, the 4-h lane is slightly overloaded, and the 6-h lane, underloaded. Nevertheless, the change in the 64-62-kD protein complex is still evident in the latter. The results illustrated here are representative of five independent experiments. (*B*) Metabolically radiolabeled methylated proteins (300,000 cpm) were obtained as described above and loaded on 8.5-15% polyacrylamide gradient gels to more favorably resolve the area between 29 and 36 kD. The image shown is a fluorogram derived from the dried gel after exposure to x-ray film for 2 d at −70°C. The arrow indicates the migration position of a regulated 34-kD protein. Cultures were treated with NGF (50 ng/ml), EGF (5 nM), or insulin (1 μM). The results shown are representative of three independent experiments.
[View Larger Versions of these Images (92 + 47K GIF file)]

To determine if any of the changes in methyl group incorporation persist for several days, the labeling studies were repeated in PC12 cells treated with NGF for 4 d. This time point was chosen because ~50% of PC12 cells have neurites ≥20-μm long after 4 d of NGF (Greene et al., 1982). In addition, it was reasoned that some protein methylations may occur during neurite outgrowth that were not detected during early stages of NGF action. Thus, after 4 d of NGF, methylated proteins were analyzed by metabolic radiolabeling of cell proteins followed by 2D IEF × SDS-PAGE (Fig. 5) as described above. The pattern of [^{3}H]methyl incorporation into PC12 cell proteins relative to total methylated protein indicates that many of the changes induced by 6 h of NGF (Fig. 1) persist for several days. Moreover, several of the NGF-induced increases and decreases in protein methylation observed after 6 h of NGF treatment were greater after 4 d of NGF treatment. Methylation of proteins between 68 and 60 kD (C, F, G, N, and N′) and in the 55-50 kD region (2-4, 6, and 7) increased relative to both the control (Fig. 5) and the 6 h NGF result (Fig. 1). Proteins C, F, G, N, and N′ appear to be the same proteins (based on pI and M_r) that showed increases in methylation after 6 h of NGF treatment. Also, after 4 d of NGF treatment, the methylation of proteins B, D, and E, as well as protein 1 in the 50-kD region decreased relative to the control or 6-h NGF result. A

number of novel methylated proteins were also observed after 4 d of NGF treatment that were not detected after 6 h. These include proteins B′, C′ in the 60-kD region of the gel, and proteins 7′, 7″, 8, and 9, appearing in the 50-kD region of the gel. The distinctly different time courses for the methylations of specific proteins suggest that NGF may regulate the methylation of proteins involved in multiple signaling pathways. Moreover, most of the NGF- dependent changes in methylated proteins that were detected are not associated with transient phenomena.

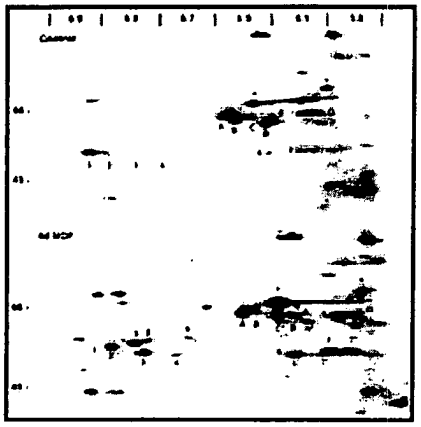


Fig. 5. NGF-induced changes in protein methylation persist and progress during prolonged NGF treatment. PC12 cells were incubated with or without NGF (50 ng/ml) for 4 d. Cells were labeled with L-[methyl-^{3}H]methionine (100 μCi/ml) in the presence of anisomycin (10 μM) for 6 h. Equal TCA-precipitable cpm (750,000) of cell lysates were subjected to IEF. IEF gels were loaded on 7.5-15% gradient SDS-PAGE gels to separate methylated proteins by molecular mass. The images shown are fluorograms of the dried gels after exposure of x-ray film for 12 d at −70°C with an intensifying screen. The areas of the gels shown are similar to the boxed areas shown in Fig. 1. The pH gradient of the IEF gels is indicated at the top of the figure. Labels are as in Fig. 1.
[View Larger Version of this Image (90K GIF file)]

Inhibition of Methylation Inhibits Neurite Outgrowth

The marked effects of NGF on the methylation of specific proteins calls attention to the possibility that protein methylation is required for NGF signal transduction and neurite outgrowth. This possibility was examined by observing NGF-mediated neurite outgrowth after inhibiting methylation using two approaches based on the biochemical pathway outlined in Fig. 6. One approach was inhibition of SAHH by a mechanism-based, substrate analogue inhibitor, DHCA (Liu et al., 1992). SAHH catalyzes the reversible hydrolysis of SAHcy to homocysteine and adenosine by an NAD-dependent mechanism (De la Haba and Cantoni, 1959; Liu et al., 1992). The reaction is driven in the hydrolysis direction in vivo by the adenosine deaminase-catalyzed conversion of adenosine to inosine. DHCA blocks the methylation pathway by inhibiting SAHH. Like the normal substrate, SAHcy, DHCA is oxidized by SAHH coupled to NAD^+ reduction, but the 3-keto DHCA formed cannot be hydrolyzed, thereby trapping SAHH in its inactive, NADH form (Liu et al., 1992). SAHH inhibition results in an accumulation of its substrate, SAHcy, a potent competitive inhibitor of methyltransferases (Hildeshein et al., 1972).

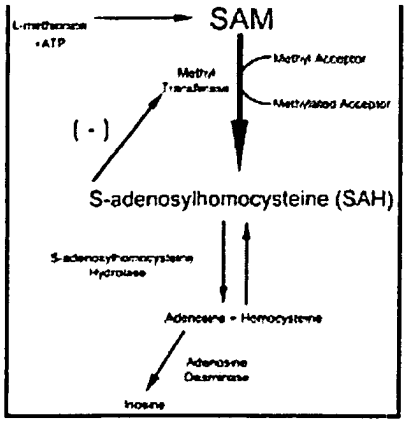


Fig. 6. The methylation pathway. SAHH catalyzes the reversible hydrolysis of SAHcy. Inhibition of SAHH by DHCA or homocysteine leads to an accumulation of SAHcy and inhibition of methyltransferases.

[View Larger Version of this Image (20K GIF file)]

The concentration dependence for inhibition of neurite outgrowth by DHCA (Fig. 7 *A*) was determined by incubating PC12 cells with NGF for 7 d with or without 10 nM to 3 μM DHCA. Neurite outgrowth was scored after 7 d of NGF treatment as described under Materials and Methods. The concentration of DHCA required to inhibit neurite outgrowth by 50% was ~100 nM. A second approach that was used to test for a requirement of protein methylation in neurite outgrowth was inhibition of SAHH by addition of extracellular homocysteine, alone or in combination with DHCA. When intracellular levels of homocysteine rise, the SAHH-catalyzed reaction is driven in the synthesis direction (De la Haba and Cantoni, 1959; Hasobe et al., 1989), i.e., towards the formation of SAHcy (Fig. 6). Excess homocysteine acts cooperatively with inhibitors of SAHH in many other cell types (Chiang et al., 1977; Kredich and Martin, 1977; Backlund et al., 1986; Hasobe et al., 1989), and the combination of the two is expected to elicit more inhibition of methylation than either alone. As shown in Fig. 7 *B*, 100 nM DHCA reduced NGF-induced neurite outgrowth to 59% of control (NGF alone), and 100 μM homocysteine reduced NGF-induced neurite outgrowth to 42%. The combination of 100 nM DHCA plus 100 μM homocysteine inhibited the neurite outgrowth by more than 80%. The findings that homocysteine alone inhibits neurite outgrowth, and that homocysteine enhances the inhibitory effect of a low dose of DHCA on neurite outgrowth indicate that the effect of DHCA on neurite outgrowth is most likely due to inhibition of methylation.

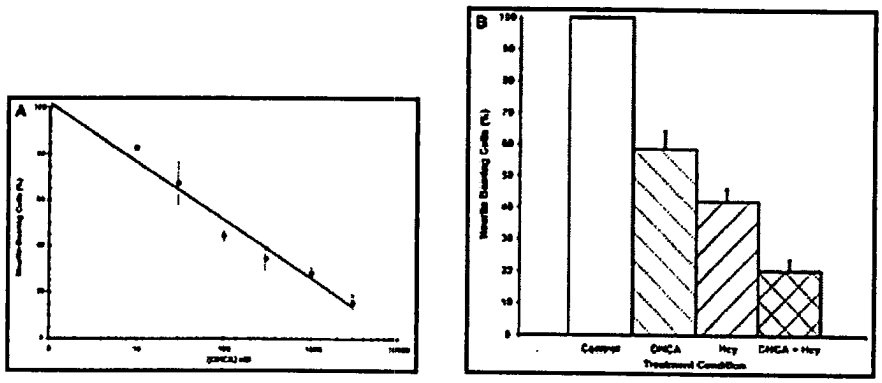


Fig. 7. Concentration-dependent inhibition of neurite outgrowth by inhibitors of methylation. (*A*) Effect of DHCA. PC12 cells on 35-mm collagen-coated tissue culture dishes (300,000 cells per dish) were treated with NGF (50 ng/ml), or NGF and DHCA (10 nM to 3 μM) for 7 d. Neurite-bearing cells were scored as described under Materials and Methods. Cultures treated with NGF and no DHCA received the DMSO

vehicle only, which had no effect on neurite outgrowth. The line was fit by least squares regression analysis. The data shown are the means ± SD for three independent experiments. (*B*) Effect of homocysteine alone and in combination with DHCA. All cultures were treated with NGF (50 ng/ml) and scored for the presence of neurites after 7 d. The data shown are the means ± SD for three independent experiments. The white bar represents the NGF control condition without DHCA or homocysteine. Results from cultures treated with 100 nM DHCA are represented by the adjacent striped bar (*DHCA*), and those from cultures treated with 100 μM homocysteine in the next striped bar (*Hcy*). The cross-hatched bar indicates results from cultures treated with 100 nM DHCA plus 100 μM homocysteine (*DHCA + Hcy*).
[View Larger Versions of these Images (8 + 18K GIF file)]

Fig. 8 illustrates the effects of DHCA or DHCA plus homocysteine on cell morphology observed by phase contrast microscopy of NGF-treated PC12 cells. After 7 d of NGF-treatment, PC12 cells flatten, hypertrophy, and elaborate long neurites (Fig. 8 *B*). In the presence of 1 μM DHCA and NGF for 7 d, the cells continue to respond to NGF by flattening, but neurite outgrowth is greatly attenuated (Fig. 8 *C*). Short, spike-like processes are observed. In the presence of 100 nM DHCA plus 100 μM homocysteine, the cells appear more flattened than control cells, but little neurite outgrowth is observed (Fig. 8 *D*). Thus, inhibition of methylation by DHCA, homocysteine, or the cooperative action of the two together suggest that methylation is an important requirement for NGF-induced neurite outgrowth. In addition, the morphological evidence (Fig. 8) indicates that not all NGF actions (e.g., the cell-flattening response) are inhibited by factors that inhibit methylation (see below).

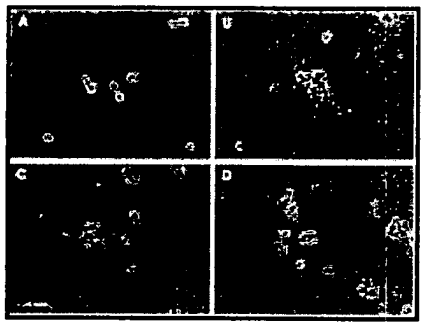


Fig. 8. The effects of DHCA and homocysteine on neurite outgrowth and cell morphology. PC12 cells were plated on 35-mm collagen-coated tissue culture dishes at a density of 300,000 cells per dish. Cells were cultured for 7 d under control conditions without NGF (*A*), with 50 ng/ml NGF (*B*), NGF plus 1 μM DHCA (*C*), or NGF with 100 nM DHCA and 100 μM homocysteine (*D*). Bar, 50 μm.

[View Larger Version of this Image (101K GIF file)]

DHCA-induced inhibition of SAHH was verified by direct measurement of SAHH activity. DHCA at 1 nM produced 70% inhibition and >90% inhibition was achieved at 10 nM. The consequence of this inhibition (elevated intracellular level of SAHcy) is expected to inhibit protein methylation (Fig. 6). Thus, the effect of DHCA on protein methylation was directly determined by assessing protein methylation in PC12 cells using metabolic radiolabeling with L-[methyl-^{3}H]methionine in the presence of anisomycin. The dose-dependent inhibition of protein methylation observed is similar to that of the DHCA effect on neurite outgrowth (Fig. 9). DHCA (1 μM) inhibits total methyl group incorporation measured by TCA-precipitation of equal aliquots of whole cell lysates as described under Materials and Methods. The total reduction in radiolabeled, methylated protein was 40% in control, non-NGF-treated cells, and 52% in cells stimulated with NGF for 6 h. Similar differences in the inhibitory effects (±NGF) were observed at all doses of DHCA above 30 nM (data not shown). The inhibitory action of DHCA is thus greater in NGF-treated cells. In conclusion, these experiments indicate that protein methylation is markedly inhibited

by DHCA at concentrations that inhibit neurite outgrowth.

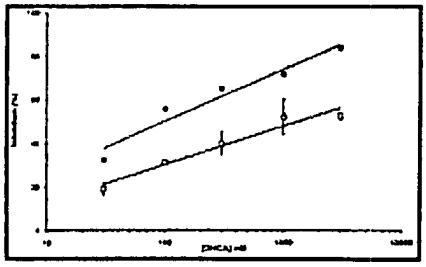


Fig. 9. Dose-dependent effect of DHCA on total protein methylation in PC12 cells correlates with the DHCA effect on inhibition of neurite outgrowth. Inhibition of NGF-stimulated neurite outgrowth (■) by DHCA was compared to the inhibition of total protein methylation (○) at the same doses of DHCA. PC12 cells were treated with DHCA (30 nM to 3 μM) and NGF (50 ng/ ml) for 6 h in the presence of anisomycin to determine total protein methylation as described under Materials and Methods. Inhibition is plotted relative to control cultures treated with NGF only. The neurite outgrowth data are replotted from Fig. 7 *A*.
[View Larger Version of this Image (8K GIF file)]

In contrast, neither NGF alone nor DHCA with or without NGF have a significant effect on total methylation of cytosines in PC12 cell DNA. The percentage of total cytosines methylated in NGF-treated cells (53.3 ± 1.5) versus that in cells treated with NGF plus DHCA (48.1 ± 0.2) was not statistically significant ($n = 3$, Student's paired t test; $P > 0.05$, two tails). All measurements were performed after 24-h treatments.

DHCA Treatment Does Not Interfere with Cell Proliferation or Early NGF-mediated Signaling Events

Previous efforts to examine the possible role of methylation in NGF signal transduction (Seely et al., 1984; Haklai et al., 1993; Kujubu et al., 1993) used methylation inhibitors at relatively high concentrations that can produce cytotoxicity (Seeley et al., 1984) and inhibition of receptor tyrosine kinases (Meakin and Shooter, 1991; Maher 1993). The methylation inhibitor used in this study, DHCA, is much less toxic than the related adenosine analogues used previously and is effective at concentrations four orders of magnitude lower than those used previously (Seeley et al., 1984; Meakin and Shooter, 1991; Haklai et al., 1993; Kujubu et al., 1993; Maher, 1993). DHCA does not affect the growth rate of PC12 cells in serum-containing medium (Fig. 10 *A*). In the absence of serum, PC12 cells degenerate via programmed cell death. DHCA has no effect on the rate that cell death is induced (Fig. 10 *A*). In the absence of serum, NGF rescues PC12 cells from programmed cell death (Rukenstein et al., 1991). Also shown in Fig. 10 *A*, DHCA had no effect on the protection afforded by NGF. In addition to confirming the absence of cytotoxicity, these findings suggest that DHCA does not interfere with the ability of NGF to activate its tyrosine kinase receptor (Trk), which prevents programmed cell death (Kaplan and Stephens, 1994). Thus DHCA has no detectable effects on the growth or survival of PC12 cells under any of the conditions examined.

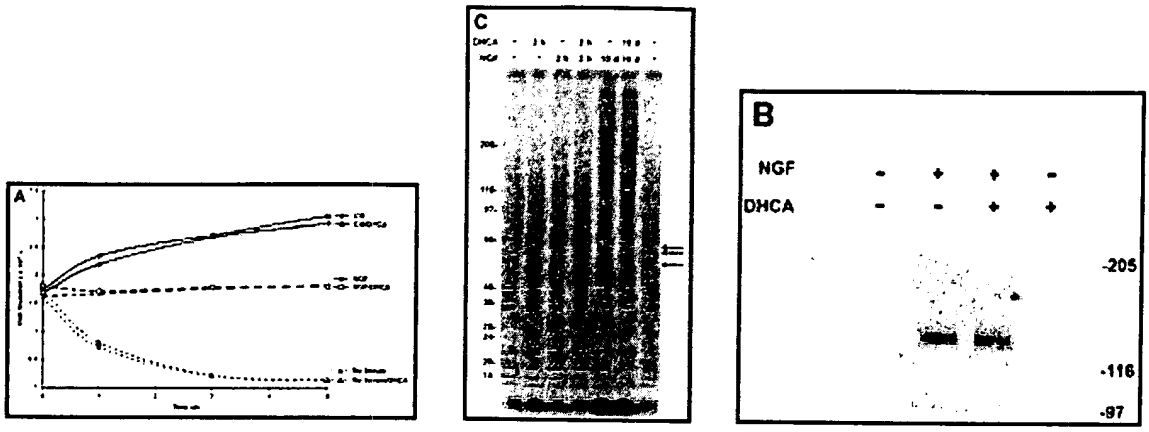


Fig. 10. DHCA treatment does not interfere with cell growth, NGF-mediated survival, or rapid onset NGF-stimulated phosphorylations. (*A*) PC12 cells were plated on 24-well collagen-coated dishes at a density of 100,000 cells per well. Cells were allowed to attach to the dishes overnight, and then were washed with RPMI-1640 medium to remove serum. Cells were then cultured in complete (serum-containing) medium (*CM*, ○ and •), plus or minus 1 μM DHCA; RPMI-1640 without serum plus 50 ng/ml NGF, plus or minus 1 μM DHCA (*NGF*, □ and ■); RPMI-1640 medium without serum or NGF, plus or minus 1 μM DHCA (*No serum*, Δ and ▲). Similar results were observed in two other independent experiments. (*B*) Effect on Trk phosphorylation. PC12 cells were treated with or without NGF (50 ng/ml) for 5 min. Where indicated, 1 μM DHCA was added to cultures for 1 h before treatment with NGF. Cell lysates were equalized for total protein, immunoprecipitated with anti-phosphotyrosine antibody (as described under Materials and Methods), resolved by 7.5% SDS-PAGE and transferred to Immobilon P membranes. The blot was probed with anti-Trk antibody, followed by donkey anti- rabbit ^{125}I-IgG. The image shown is an autoradiogram of the Western blot after a 5-d exposure at −80°C. This result was verified in a second independent experiment. (*C*) PC12 cells were treated with or without NGF (50 ng/ml) for 2 h. DHCA (1 μM) was added 1 h before the addition of NGF where indicated. Cells were radiolabeled with [^{32}P]orthophosphate for 2 h as described under Materials and Methods. For long term NGF-induced phosphorylations, PC12 cells were grown in the presence of NGF (50 ng/ml) plus or minus 1 μM DHCA for 10 d. Cells were then radiolabeled with [^{32}P]orthophosphate for 2 h. For both short and long term NGF treatment experiments, cell lysates were equalized for total TCA-precipitable ^{32}P-labeled protein and separated on 6-12% gradient SDS-PAGE. The image shown is an autoradiogram of the dried gel. Exposure time was 16 h. The arrows at the right indicate, in decreasing M_r, the positions of the previously identified phosphoproteins, 64-kD chartin, 60-kD tyrosine hydroxylase, and 55-kD β-tubulin.
[View Larger Versions of these Images (12 + 62 + 40K GIF file)]

Since other adenosine-analogues, which are putative methylation inhibitors, are known to inhibit phosphorylation and activation of Trk (Meakin and Shooter, 1991; Maher 1993), the effect of DHCA on the tyrosine phosphorylation of Trk was also determined. PC12 cells were incubated with or without 1 μM DHCA for 1 h, and NGF was then added for 5 min. Immunoprecipitation with a phosphotyrosine antibody was followed by SDS-PAGE and immunoblotting with anti-Trk antibody. DHCA had no detectable effect on the Trk tyrosine phosphorylation mediated by NGF (Fig. 10 *B*), the primary event in NGF signaling.

To determine if DHCA affects other phosphorylation events downstream of Trk activation, total

phosphoproteins were analyzed by SDS-PAGE. Several NGF-induced increases in phosphorylation of proteins are resolved by this method (Fig. 10 *C*). The rapid increase in the phosphorylation of tyrosine hydroxylase (Halegoua and Patrick, 1980) that occurs after NGF treatment is not affected by DHCA. The overall phosphorylation pattern was similarly unaffected by DHCA. Two other NGF-induced increases in phosphorylation, which require long-term (7 d) NGF treatment, are temporally associated with neurite outgrowth. Unlike phosphorylation of tyrosine hydroxylase, the increased ^{32}P incorporation into the 64-kD chartin (Aletta and Greene, 1987) and β-tubulin (Aletta, 1996; Black et al., 1986) is diminished by exposure of the cells to DHCA. The inhibitory effect of DHCA on the phosphorylations of these two proteins is the predicted consequence of DHCA inhibition of neurite outgrowth. Thus, DHCA affects delayed phosphorylation events in NGF signaling but does not inhibit phosphorylations associated with the early signaling events studied here.

DHCA Effects on Neurite Outgrowth Are Diminished by prior NGF Treatment and Are Rapidly Reversible

To explore further the functional significance of the time dependence of protein methylation during NGF-mediated neurite outgrowth, DHCA (1 μM) was added to PC12 cells at the same time that NGF was added or 1-6 d after NGF addition. Under these conditions, the later DHCA was added after NGF, the less its inhibitory effect (Fig. 11 *A*). The decrease in inhibition showed a linear dependence on the time at which DHCA was added to the cells. Once the NGF-signaling cascade begins, neurite outgrowth is progressively less sensitive to inhibition of methylation. This would occur if the rates of demethylation were low for those methylated proteins that are required for neurite outgrowth. Alternatively, neurite outgrowth may be insensitive to inhibiting methylation once neurite outgrowth proceeds past a critical, committed step. While initiation of neurite outgrowth from PC12 cells after exposure to NGF is slow (requiring 7 d to achieve ~100% neurite bearing cells), PC12 cells can rapidly regenerate after preformed neurites are disrupted as described under Materials and Methods. Fig. 11 *B* shows the concentration dependence of inhibition of neurite regeneration by DHCA. Neurite regeneration was inhibited 40% by 3 μM DHCA. In contrast, if 3 μM DHCA is used when NGF is first added to the cells, it inhibits neurite outgrowth by 84% (Fig. 7 *A*). As described above, the difference in the sensitivities of regeneration and de novo neurite outgrowth to inhibition by DHCA is consistent with slow rates of demethylation or decreased sensitivity to DHCA once differentiation has reached a critical stage.

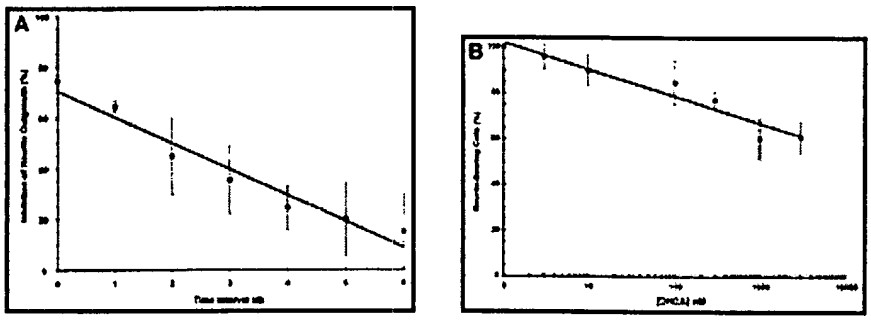


Fig. 11. The effect of prior NGF treatment on the neurite inhibiting action of DHCA. (*A*) The effect of varying the time interval between NGF treatment and addition of DHCA. PC12 cells were treated with NGF (50 ng/ml) for 7 d. DHCA (1 μM) was added to the cells either at the time NGF was initially added to the cell cultures (time zero), or after a time interval of 1, 2, 3, 4, 5, or 6 d. Cells were scored for the

presence of neurites after 7 d of NGF. The data shown are the means ± SD for three independent experiments. (*B*) The effect of inhibition of methylation on neurite regeneration. PC12 cells were grown in NGF (50 ng/ml) for 7-14 d. NGF was removed from the cultures by an extensive washing procedure described under Materials and Methods. Cells were removed from the culture dishes by trituration and replated without or with DHCA (3 nM to 3 μM). NGF was added to all cultures 1 h after plating. Neurite regeneration was scored 16-24 h later. The line was fit by least squares regression analysis. The data shown are the means ± SD for three independent experiments.
[View Larger Versions of these Images (10 + 8K GIF file)]

The onset of neurite outgrowth after DHCA is removed was examined to compare the rate of reversing the effect of DHCA on neurite outgrowth to the rate of NGF- induced neurite outgrowth when DHCA had never been present. In addition, removal of DHCA was used to detect protein methylations that are associated with neurite outgrowth during the reversal period. Reversibility of the effects of DHCA on neurite outgrowth and protein methylation was assessed by incubating PC12 cells with NGF and DHCA for 10 d, followed by removal of the DHCA. Interestingly, the inhibitory effect of DHCA on neurite outgrowth was rapidly reversed (Fig. 12 *A*). No lag was observed before an increase in neurite outgrowth was detected. The $t_{1/2}$ for reversal was 12 h (Fig. 12 *A*). The rate of neurite outgrowth when DHCA is removed is, therefore, about eightfold faster than the normal $t_{1/2}$ for neurite outgrowth when NGF is added to naive cells. Thus, the NGF-dependent signaling events that prime the cells (Burstein and Greene, 1978; Greene et al., 1982) for neurite outgrowth were not inhibited greatly by the methylation inhibitor, DHCA.

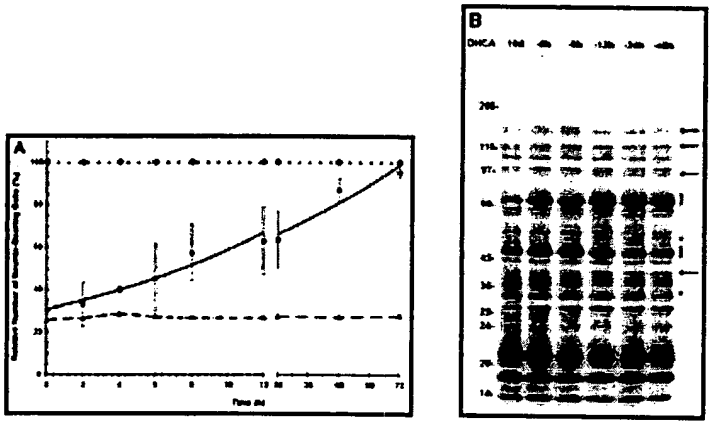


Fig. 12. Reversal of DHCA-induced inhibition of methyltransferases results in rapid neurite outgrowth and a time-dependent increase in protein methylation. (*A*) PC12 cells were plated on 35-mm collagen-coated tissue culture dishes at a density of 300,000 cells per dish in NGF (50 ng/ ml) plus or minus 1 μM DHCA for 7 d. The medium was then changed repeatedly to remove DHCA, and neurite outgrowth was scored and plotted over a 3-d period. Controls were maintained in DHCA/NGF (▲) and NGF alone (•). The experimental condition was treatment with DHCA and NGF for 7 d, followed by NGF alone for 3 d (■). The data are normalized to 100% relative to the number of neurite bearing cells present in cultures treated with NGF alone. The data shown are the means ± SD for three independent experiments. The line shown for the reversal data is the best fit to a single exponential function with a $t_{1/2}$ of 12 h. After DHCA removal, the protein methylation pattern was also assessed by SDS-PAGE. (*B*) Cells were harvested after 10 d of

DHCA/NGF treatment or after 7 d DHCA/NGF treatment followed by progressively longer times (*-h*) in the absence of DHCA. Cell lysates were equalized for total TCA-precipitable ^3H counts per minute and separated on 7.5-15% gradient SDS-PAGE. The figure shown is a fluorogram of the dried gel. Exposure time was 5 d. The arrows and brackets indicate the positions of proteins that increased in methylation by 25% or more relative to the 10 d DHCA lane. Asterisks mark the position of protein bands not detected in the 10-d DHCA lane. The changes noted during the reversal period were replicated in four independent experiments.
[View Larger Versions of these Images (13 + 76K GIF file)]

Upon removal of DHCA, the rapid onset of neurite outgrowth was paralleled by rapid reversal of its inhibitory effects on the methylation of several proteins. Proteins that increase by 25% or more in methylation upon removal of DHCA are indicated by the arrows, brackets, and asterisks shown in Fig. 12 *B*. The molecular weights of proteins that showed increased methylation migrate at 150, 120, 94, 70/68, 50, 48/47, 40, and 35 kD. The time dependencies for methylation were different for different proteins. These results, together with the time course (Figs. 4 and 5), inhibitor studies (Figs. 7-10), and specificity of the NGF- induced effects on protein methylation (Fig. 1) indicate that regulation of protein methylation is required for neurite outgrowth.

Discussion

The Subset of Methylated Proteins that Are NGF-regulated

This work extends previous studies (Seeley et al., 1984; Kujubu et al., 1993) of the role of methylation in growth factor signaling mechanisms and further demonstrates the feasibility of exploring this role by metabolic radiolabeling of specific proteins in intact cells and by in vitro labeling of proteins in cell fractions. The results indicate that protein methylation is involved in the NGF-induced signal transduction that leads to neurite outgrowth. Methylation of numerous proteins was detected by metabolic radiolabeling indicating that protein methylation is an ongoing, active process. NGF specifically affects the methylation of a subset of these proteins. Further evidence for a significant role for protein methylation in neurite outgrowth is the marked concurrent inhibitory effects of DHCA on protein methylation and neurite outgrowth. Moreover, removal of DHCA rapidly leads to concomitant increases in protein methylation and neurite outgrowth. The identities and functions of the NGF-regulated proteins remain to be determined. It is likely that some of the proteins identified by this approach represent previously unidentified players in NGF signal transduction. This is significant because specific protein modifications that are temporally coincident with neuronal differentiation may be useful in the elucidation of the mechanism of action of NGF (Chao, 1992; Kaplan and Stephens, 1994). Thus, signal transduction pathways may be regulated in part by regulating the methylation of specific proteins in much the same manner as protein phosphorylation regulates signaling pathways.

Potential Actions of Protein Methylation During NGF Signal Transduction

The data presented here focus on NGF actions on protein methylations rather than methylation of

phospholipids, RNA or DNA. Methylation of genes serves as a general signal for repression of transcription (Razin and Cedar, 1991). Hypomethylation of the genome is associated with cellular differentiation in vitro (Bestor et al., 1984). There were no significant effects of NGF and/or DHCA on total DNA cytosine methylation in PC12 cells. Thus, the potent inhibitory action of DHCA on neurite outgrowth is unlikely to be due to an effect on DNA methylation. Previous work also indicates that NGF does not detectably affect phospholipid methylation (Ferrari and Greene, 1985). These authors further demonstrated that complete inhibition of phospholipid methylation is achieved at concentrations of deaza-adenosine and homocysteine thiolactone that do not block neurite outgrowth.

There are several potential mechanisms through which NGF-induced changes in protein methylation may affect biological responses. Gene regulation may be affected, particularly if the modified proteins are involved in gene expression. For example, many heterogeneous nuclear RNPs that associate with pre-mRNAs are methylated on arginine residues (Liu and Dreyfuss, 1995). Modulation of RNA splicing has been proposed as one potential target for protein methylation in signal transduction (Lin et al., 1996). Consistent with this view, a protein arginine methyltransferase activity appears to be constitutively associated with the type I interferon receptor (Abramovich et al., 1997). A role for protein methylation in cell signaling is further substantiated by the finding that methylation of the γ subunit of a heterotrimeric G-protein enhances activation of PI-specific phospholipase C and PI3-kinase in vitro (Parish et al., 1995). If this biochemical effect occurs in intact cells, protein methylation could play an important general role in the amplification of intracellular signals. Protein methylation may also be involved in altering the subcellular localization of specific proteins, as recently suggested for the cell cycle-dependent methylation of protein phosphatase 2A (Turowski et al., 1995).

Neurite outgrowth requires a fundamental reorganization of the cytoskeleton including formation of intermediate filaments and parallel bundles of microtubules oriented longitudinally in neurites, as well as specialized actin microfilament arrays within motile growth cones. Inhibition of neurite outgrowth by DHCA may be related to inhibition of protein methylation, which affects the cytoskeleton. For example, Ras and Rho proteins, modified by isoprenylation and methylation, are involved in redistribution of actin microfilaments during changes in cell shape (Vojtek and Cooper, 1995). Recently, Luo et al. (1996) presented evidence that constitutively active Rac1 in transgenic mice interferes with the normal formation of Purkinje cell axons and dendrites. Since only the methylation step of small G-protein processing is reversible (Rando, 1996), methylation is more likely than other processing steps to be subject to physiological regulation. NGF-induced increased methyl labeling of G-proteins in PC12 cells (Haklai et al., 1993; Kujubu et al., 1993) and the specific proteins observed in this study indicate that NGF may serve as one mediator of regulatory control in responsive neurons. Furthermore, Klein and colleagues have raised the possibility that protein methylation plays an important role in neurulation (Coelho and Klein, 1990; Moephuli et al., 1997) and neurite outgrowth in rat embryo cultures (Coelho and Klein, 1990). The present work lends additional support to this possibility, especially with regard to neurite outgrowth.

DHCA Is a Potent and Nontoxic Inhibitor of Protein Methylation

DHCA is a promising tool for investigating the possible roles of methylation in signal transduction mechanisms. This inhibitor was developed by Liu and colleagues to obtain a more specific, less toxic

inhibitor of SAHH than previously used inhibitors (Liu et al., 1992). The 50% effective dosage (ED_{50}) for the effect of DHCA on neurite outgrowth for PC12 cells (~100 nM) is four orders of magnitude lower than previously used inhibitors (Seeley et al., 1984). Inhibition of neurite outgrowth by >~85% is difficult to achieve, even at the highest concentration of the inhibitor. Pretreatment of cells with DHCA for several weeks before NGF stimulation, however, does produce nearly complete inhibition even at 7 d of NGF (Cimato, T.R., M.J. Ettinger, and J.M. Aletta, unpublished observations). These observations provisionally indicate that basal, ongoing protein methylation may serve an auxiliary role in neurite outgrowth in addition to the NGF effects on protein methylation. As expected, DHCA potently inhibits SAHH activity (Hasobe et al., 1989) and consequently leads to decreased total protein methylation. The dose-dependent inhibition of protein methylation observed is quite similar to that for inhibition of neurite outgrowth (Fig. 9). The cooperative effects of homocysteine and DHCA on neurite outgrowth further substantiate a specific effect on methylation. The finding that total protein methylation was inhibited more when NGF was present than when absent, suggests that under the influence of NGF, many of the potential methyltransferases involved in the NGF response may be more sensitive to inhibition due to increased levels of SAHcy. This may occur because, as a result of increased total protein methylation (Fig. 3), intracellular levels of SAHcy are expected to be elevated even before DHCA treatment. NGF has no measureable effect on SAHH activity (Cimato, T.R., M.J. Ettinger, and J.M. Aletta, unpublished results).

DHCA is unique among methylation inhibitors because of its lack of effect on cell flattening, phosphorylation of tyrosine hydroxylase, and tyrosine phosphorylation of Trk. Less specific methylation inhibitors, used to study the role of methylation in PC12 cells previously (Seeley et al., 1986; Kujubu et al., 1993; Maher, 1993; Meakin and Shooter, 1995), blocked each of these actions. The most remarkable property of DHCA in comparison to previously used inhibitors is that DHCA inhibits neurite outgrowth without inhibiting the NGF priming that prepares PC12 cells for neurite outgrowth (Greene et al., 1982). This is best illustrated by the rapid reversal of DHCA effects on neurite outgrowth, which is in marked contrast to that observed with other methylation inhibitors. Reversal of inhibition by 5′-deoxy-S-methyl adenosine requires the same 7 d as required by control PC12 cells when first exposed to NGF (Seeley et al., 1986). Thus, DHCA is clearly more specific than previously used methylation inhibitors. DHCA specifically affects neurite outgrowth without affecting many other signaling effects of NGF, including those associated with priming.

Protein Methylation in Early and Delayed NGF Signaling

Growth factors in general, promote biological responses that are either extremely rapid (within minutes) or require many hours or days to develop fully. The diversity of signaling networks (Pawson, 1985; Hunter, 1995) and the propensity of specific pathways to transmit and/or sustain a given signal (Marshall, 1995) may help to explain the different temporal characteristics of responses. The results of this study and other recent work (Metz et al., 1993; Philips et al., 1993, 1995) indicate that posttranslational modification of proteins by methylation offers another potential mechanism for regulating both early and delayed growth factor signaling.

The importance of early and delayed responses to NGF in PC12 cells, dorsal root ganglion, and sympathetic neurons is well documented (Greene, 1984). One of the most obvious examples of a delayed

NGF response is neurite outgrowth. Little neurite outgrowth is evident in PC12 cells treated with NGF for 12-24 h. 50% of the cells extend neurites within 4 d, and nearly all have neurites by 7 d. NGF-mediated neurite outgrowth requires transcription (Burstein and Greene, 1978; Greene et al., 1982). The long latency is apparently due to the progressive accumulation of specific proteins required for neurite assembly during the priming period. This accumulation may depend on protein synthesis (Burstein and Greene, 1978) and/or progressive posttranslational modifications of specific proteins (e.g., as shown here for the methylation of specific proteins). Since regeneration of neurites from PC12 cells previously treated with NGF occurs much more rapidly (within 24 h) than initiation (7 d); the longer time required for neurite outgrowth in the latter is thought to be due to the "priming" events.

When interpreted in light of the priming model of NGF-induced neurite outgrowth, the experiments that used DHCA provide important insights regarding the temporal nature of protein methylation involved in this NGF response. Inhibition of protein methylation by DHCA is most effective in reducing neurite formation when the DHCA is present concomitantly with NGF treatment. There is decreased inhibition of neurite outgrowth as the time interval between NGF treatment and the addition of DHCA increases (Fig. 11 *A*). DHCA is also less effective at inhibiting neurite regeneration (Fig. 11 *B*). These results imply progressive accumulation of the methylated protein product(s) necessary for neurite outgrowth, as was detected at 4 d of NGF treatment (Fig. 5). Slow turnover of methylated proteins is consistent with the reduced effect of DHCA when added after NGF signaling has been initiated.

The time dependencies of specific protein methylations, the effects of DHCA on protein methylation, and the lack of effects of DHCA on early protein phosphorylations and priming are consistent with little or no interdependence between protein methylation signaling events and early, priming events. Moreover, the data presented on the phosphorylation of Trk and other proteins that lie downstream of Trk activation (Fig. 10, *B* and *C*) further substantiate the view that some cellular signaling events are more susceptible to interference from inhibition of protein methylation than others. Autophosphorylation of Trk and the rapid increase in the phosphorylation of tyrosine hydroxylase (Halegoua and Patrick, 1980) are not affected by DHCA. Two other NGF-induced increases in phosphorylation, which require long term (≥7 d) NGF treatment for maximal phosphorylation, are temporally associated with neurite outgrowth. In accord with DHCA-mediated inhibition of neurite outgrowth, but unlike the rapid, early phosphorylation events, the increased ^{32}P incorporation into the 64-kD chartin (Aletta and Greene, 1987) and β-tubulin (Black et al., 1986; Aletta, 1996) is diminished by exposure of the cells to DHCA.

It can thus be concluded that DHCA does not disrupt all early NGF signaling events and priming actions, nor does it abrogate many of the biological effects of NGF signaling (e.g., Fig. 10 *A*, survival; Fig. 9, cell flattening). The effects of DHCA on neurite outgrowth and inhibition of the phosphorylation of neurite-associated proteins (β-tubulin and chartin) are consistent with a selective inhibitory action on an NGF pathway responsible for neurite outgrowth. Based on all of the above results, the protein methylation required is likely to occur downstream from, or parallel to, rapid NGF actions and priming events. This interpretation does not exclude the possibility that protein methylation of rapid onset also plays an important role in neurite outgrowth. Once priming is achieved, NGF stimulation of some of the rapid protein methylations observed in this study may assume added significance, particularly if they occur at the growth cone.

Footnotes

Received for publication 20 March 1997 and in revised form 17 June 1997.

Please address all correspondence to John M. Aletta, Department of Pharmacology and Toxicology, University at Buffalo School of Medicine and Biomedical Sciences, 3435 Main Street, Buffalo, NY 14214-3000. Tel.: (716) 829-3237. Fax: (716) 829-2801. e-mail: jaletta@ubmed.buffalo.edu

We are grateful to R. Borchardt for synthesizing the DHCA that was used for these studies.

This work was supported in part by USDA grant No. 9500773, and by the Multidisciplinary Research Pilot Program from the University at Buffalo, State University of New York.

Abbreviations used in this paper

DHCA, 9-(*trans*-2′, *trans*-3′-dihydroxycyclopent-4′-enyl)-adenine; ERK, extracellular signal-regulated kinase; PAS, protein-A-Sepharose; 2D two dimensional; PI, phosphoinositide; SAHcy, S-adenosylhomocysteine; SAHH, S-adenosylhomocysteine hydrolase; SAM, S-adenosylmethionine; Trk, tyrosine kinase receptor.

References

1. Abramovich, C., B. Yakobson, J. Chebath, and M. Revel. 1997. A protein-arginine methyltransferase binds to the intracytoplasmic domain of the IFNAR1 chain in the type I interferon receptor. *EMBO (Eur. Mol. Biol. Organ.) J.* 16: 260-266 [Abstract/Full Text].
2. Aletta, J.M.. 1996. Phosphorylation of type III β-tubulin in PC12 cell neurites during NGF-induced process outgrowth. *J. Neurobiol.* 31: 461-475 [Medline].
3. Aletta, J.M., and L.A. Greene. 1987. Sequential phosphorylation of chartin microtubule-associated proteins is regulated by the presence of microtubules. *J. Cell Biol.* 105: 277-290 [Abstract].
4. Backlund, P.S. Jr., D. Carotti, and G.L. Cantoni. 1986. Effects of the S-adenosylhomocysteine hydrolase inhibitors 3-deazaadenosine and 3-deazaaristeromycin on RNA methylation and synthesis. *Eur. J. Biochem.* 160: 245-251 [Abstract].
5. Bestor, T.H., S.B. Hellewell, and V.M. Ingram. 1984. Differentiation of two mouse cell lines is associated with hypomethylation of their genomes. *Mol. Cell. Biol.* 4: 1800-1806 [Medline].
6. Black, M.M., J.M. Aletta, and L.A. Greene. 1986. Regulation of microtubule composition and stability during nerve growth factor-promoted neurite outgrowth. *J. Cell Biol.* 103: 545-557 [Abstract].
7. Bonner, W.M., and R.A. Laskey. 1974. A film detection method for tritium- labeled proteins and nucleic acids in polyacrylamide gels. *Eur. J. Biochem.* 46: 83-88 [Medline].
8. Bradford, M.M.. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72: 248-254 [Medline].
9. Burstein, D.E., and L.A. Greene. 1978. Evidence for RNA synthesis-dependent and independent

pathways in stimulation of neurite outgrowth by nerve growth factor. *Proc. Natl. Acad. Sci. USA.* 75: 6059-6063 [Medline].

10. Chamberlain, J.P.. 1979. Fluorographic detection of radioactivity in polyacrylamide gels with the water-soluble fluor, sodium salicylate. *Anal. Biochem.* 98: 132-135 [Medline].
11. Chao, M.V.. 1992. Growth factor signaling: where is the specificity? *Cell.* 68: 995-997 [Medline].
12. Chelsky, D., B. Ruskin, and D.E. Koshland. 1985. Methyl-esterified proteins in a mammalian cell line. *Biochemistry.* 24: 6651-6658 [Medline].
13. Chiang, P.K., H.H. Richards, and G.L. Cantoni. 1977. S-adenosyl-L-homocysteine hydrolase: analogues of S-adenosyl-L-homocysteine as potential inhibitors. *Mol. Pharmacol.* 13: 939-947 [Abstract].
14. Coelho, C.N.D., and N.W. Klein. 1990. Methionine and neural tube closure in cultured rat embryos: morphological and biochemical analyses. *Teratology.* 42: 437-451 [Medline].
15. De la Haba, G., and G.L. Cantoni. 1959. The enzymatic synthesis of S-adenosyl-L-homocysteine from adenosine and homocysteine. *J. Biol. Chem.* 234: 603-608 .
16. Favre, B., S. Zolnierowicz, P. Turowski, and B.A. Hemmings. 1994. The catalytic subunit of protein phosphatase 2A is carboxyl-methylated in vivo. *J. Biol. Chem.* 269: 16311-16317 [Abstract].
17. Ferrari, G., and L.A. Greene. 1985. Does phospholipid methylation play a role in the primary mechanism of action of nerve growth factor? *J. Neurochem.* 45: 853-859 [Abstract].
18. Greene, L.A.. 1984. The importance of both early and delayed responses in the biological actions of nerve growth factor. *Trends Neurosci.* 13: 91-94 .
19. Greene, L.A., and A.S. Tischler. 1976. Establishment of a noradrenergic clonal line of rat adrenal pheochromocytoma cells which respond to nerve growth factor. *Proc. Natl. Acad. Sci. USA.* 73: 2424-2428 [Medline].
20. Greene, L.A, D.E. Burstein, and M.M. Black. 1982. The role of transcription-dependent priming in nerve growth factor promoted neurite outgrowth. *Dev. Biol.* 91: 305-316 [Medline].
21. Greene, L.A., J.M. Aletta, A. Rukenstein, and S.H. Green. 1987. PC12 pheochromocytoma cells: culture, nerve growth factor treatment, and experimental exploitation. *Methods Enzymol.* 147: 207-216 [Medline].
22. Haklai, R., S. Lerner, and Y. Kloog. 1993. Nerve growth factor induces a succession of increases in isoprenylated methylated small GTP-binding proteins of PC12-pheochromocytoma cells. *Neuropeptides.* 24: 11-25 [Medline].
23. Halegoua, S., and J. Patrick. 1980. Nerve growth factor mediates phosphorylation of specific proteins. *Cell.* 22: 571-581 [Medline].
24. Hasobe, M., J.G. McKee, H. Ishii, M. Cools, R.T. Borchardt, and E. DeClercq. 1989. Elucidation of the mechanism by which homocysteine potentiates the anti-vaccinia virus effects of the S-adenosylhomocysteine hydrolase inhibitor 9-(trans-2′, trans-3′-dihydroxycyclopent-4′-enyl) derivatives of adenine and 3-deazaadenine on the metabolism of S-adenosylhomocysteine in mouse L929 cells. *Mol. Pharmacol.* 36: 490-496 [Abstract].
25. Hershfield, M.S.. 1979. Apparent suicide inactivation of human lymphoblast S-adenosylhomocysteine hydrolase by 2′-deoxyadenosine and adenosine arabinoside. *J. Biol. Chem.* 254: 22-25 [Medline].
26. Hildeshein, J., R. Hildeshein, and E. Lederer. 1972. New syntheses of S-adenosyl homocysteine analogues, potential methyltransferase inhibitors. *Biochimie (Paris).* 54: 989-995 [Medline].
27. Hrycyna, C.A., and S. Clarke. 1993. Modification of eukaryotic signaling proteins by COOH-terminal methylation reactions. *Pharmacol. Ther.* 59: 281-300 [Medline].
28. Huff, K., D. End, and G. Guroff. 1981. Nerve growth factor-induced alteration in the response of PC12 pheochromocytoma cells to epidermal growth factor. *J. Cell Biol.* 88: 189-198 [Medline].

29. Hunter, T.. 1995. Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling. *Cell.* 80: 225-236 [Medline].

30. Kaplan, D.R., and R.M. Stephens. 1994. Neurotrophin signal transduction by the Trk receptor. *J. Neurobiol.* 25: 1404-1417 [Medline].

31. Kredich, N.M., and D.W. Martin Jr.. 1977. Role of S-adenosylhomocysteine in adenosine mediated toxicity in cultured mouse T lymphoma cells. *Cell.* 12: 931-938 [Medline].

32. Kujubu, D.A., J.B. Stimmel, R.E. Law, H.R. Herschman, and S. Clarke. 1993. Early responses of PC12 cell to NGF and EGF: effect of K252a and 5′-methylthioadenosine on gene expression and membrane protein methylation. *J. Neurosci. Res.* 36: 58-65 [Medline].

33. Laemmli, U.K.. 1970. Cleavage of structural proteins during assembly of the head of bacteriophage T4. *Nature (Lond.).* 222: 680-685 .

34. Laskey, R.A., and A.D. Mills. 1975. Quantitative film detection of ^{3}H and ^{14}C in polyacrylamide gels by fluorography. *Eur. J. Biochem.* 56: 335-341 [Abstract].

35. Lee, J., and J. Stock. 1993. Protein phosphatase 2A catalytic subunit is methyl-esterified at its carboxyl terminus by a novel methyltransferase. *J. Biol. Chem.* 268: 19192-19195 [Abstract].

36. Lee, J., Y. Chen, T. Tolstykh, and J. Stock. 1996. A specific protein carboxyl methylesterase that demethylates phosphoprotein phosphatase 2A in bovine brain. *Proc. Natl. Acad. Sci. USA.* 93: 6043-6047 [Abstract].

37. Lin, W.-J., J.D. Gary, M.C. Yang, S. Clarke, and H.R. Herschman. 1996. The mammalian immediate-early gene TIS21 protein and the leukemia-associated BTG1 protein interact with a protein-arginine *N*-methyltransferase. *J. Biol. Chem.* 271: 15034-15044 [Abstract/Full Text].

38. Liu, Q., and G. Dreyfuss. 1995. In vivo and in vitro arginine methylation of RNA-binding proteins. *Mol. Cell. Biol.* 15: 2800-2808 [Abstract].

39. Liu, S., M.S. Wolfe, and R.T. Borchardt. 1992. Rational approaches to the design of antiviral agents based on S-adenosyl-L-homocysteine hydrolase as a molecular target. *Antivir. Res.* 19: 247-265 [Medline].

40. Luo, L., T.K. Hensch, L. Ackerman, S. Barbel, L.Y. Jan, and Y.N. Jan. 1996. Differential effects of the Rac GTPase on Purkinje cell axons and dendritic trunks and spines. *Nature (Lond.).* 379: 837-840 [Medline].

41. Maher, P.A.. 1993. Inhibition of the tyrosine kinase activity of fibroblast growth factor receptor by the methyltransferase inhibitor 5′-methylthioadenosine. *J. Biol. Chem.* 268: 4244-4249 [Abstract].

42. Marshall, C.J.. 1995. Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation. *Cell.* 80: 179-185 [Medline].

43. Meakin, S.O., and E.M. Shooter. 1991. Tyrosine kinase activity coupled to the high-affinity nerve growth factor-receptor complex. *Proc. Natl. Acad. Sci. USA.* 88: 5862-5866 [Abstract].

44. Metz, S.A., M.E. Rubaglia, J.B. Stock, and A. Kowluru. 1993. Modulation of insulin secretion from normal rat islets by inhibitors of the post-translational modifications of GTP binding proteins. *Biochem. J.* 295: 31-40 [Medline].

45. Miller, M.W., D.M.J. Duhl, B.M. Winkes, F. Arredondo-Vega, P.J. Saxon, G.L. Wolff, C.J. Epstein, M.S. Hershfield, and G.S. Barsh. 1994. The mouse *lethal nonagouti* (a^{x}) mutation deletes the *S-adenosylhomocysteine hydrolase (Ahcy)* gene. *EMBO (Eur. Mol. Biol. Organ.) J.* 13: 1806-1816 [Abstract].

46. Mobley, W.C., A. Schenker, and E.M. Shooter. 1976. Characterization and isolation of proteolytically modified nerve growth factor. *Biochemistry.* 15: 5543-5552 [Medline].

47. Moephuli, S.R., N.W. Klein, M.T. Baldwin, and H.M. Krider. 1997. Effects of methionine on the cytoplasmic distribution of actin and tubulin during neural tube closure in rat embryos. *Proc. Natl.*

Acad. Sci. USA. 94: 543-548 [Abstract/Full Text].

48. Najbauer, J., and D.W. Aswad. 1990. Diversity of methyl acceptor proteins in rat pheochromocytoma (PC12) cells revealed after treatment with adenosine dialdehyde. *J. Biol. Chem.* 265: 12717-12721 [Abstract].
49. Parish, C.A., A.V. Smrcka, and R.R. Rando. 1995. Functional significance of $\beta\gamma$-subunit carboxyl methylation for the activation of phospholipase C and phosphoinositide 3-kinase. *Biochemistry.* 34: 7722-7727 [Medline].
50. Pawson, T.. 1985. Protein modules and signaling networks. *Nature (Lond.).* 373: 573-580 .
51. Peng, X., L.A. Greene, D.R. Kaplan, and R.M. Stephens. 1995. Deletion of a conserved juxtamembrane sequence in trk abolishes NGF-promoted neuritogenesis. *Neuron.* 15: 395-406 [Medline].
52. Philips, M.R., M.H. Pillinger, R. Staud, C. Volker, M.G. Rosenfeld, G. Weissmann, and J.B. Stock. 1993. Carboxyl methylation of Ras-related proteins during signal transduction in neutrophils. *Science (Wash. DC).* 259: 868-872 .
53. Philips, M.R., R. Staud, M. Pillinger, A. Feoktistov, C. Volker, J.B. Stock, and G. Weissmann. 1995. Activation-dependent carboxyl methylation of neutrophil G-protein γ subunit. *Proc. Natl. Acad Sci. USA.* 92: 2283-2287 [Abstract].
54. Qui, M.-S., and S.H. Green. 1992. PC12 cell neuronal differentiation is associated with prolonged $p21^{ras}$ activity and consequent prolonged ERK activity. *Neuron.* 9: 705-717 [Medline].
55. Rando, R.R.. 1996. Chemical biology of protein isoprenylation/methylation. *Biochim. Biophys. Acta.* 1300: 5-16 [Medline].
56. Razin, A., and H. Cedar. 1991. DNA methylation and gene expression. *Microbiol. Rev.* 55: 451-458 [Medline].
57. Rukenstein, A., R.E. Rydel, and L.A. Greene. 1991. Multiple agents rescue PC12 cells from serum-free cell death by translation, and transcription- dependent mechanisms. *J. Neurosci.* 11: 2557-2563 .
58. Seeley, P.J., A. Rukenstein, J.L. Connolly, and L.A. Greene. 1984. Differential inhibition of nerve growth factor and epidermal growth factor on the PC12 pheochromocytoma line. *J. Cell Biol.* 98: 417-426 [Abstract].
59. Shapiro, M.J., I. Chakrabarti, and D.E. Koshland. 1995. Contributions made by individual methylation sites of *Escherichia coli* aspartate receptor to chemotactic behavior. *Proc. Natl. Acad. Sci. USA.* 92: 1053-1056 [Abstract].
60. Soto, A.M., and C. Sonnenschein. 1985. The role of estrogens on the proliferation of human breast tumor cells (MCF-7). *J. Steroid Biochem.* 23: 87-94 [Medline].
61. Tischler, A.S., and L.A. Greene. 1978. Morphologic and cytochemical properties of a clonal line of rat adrenal pheochromocytoma cells which respond to nerve growth factor. *Lab. Invest.* 39: 77-89 [Medline].
62. Turowski, P., A. Fernandez, B. Favre, N.J.C. Lamb, and B.A. Hemmings. 1995. Differential methylation and altered conformation of cytoplasmic and nuclear forms of protein phosphatase 2A during cell cycle progression. *J. Cell Biol.* 129: 397-410 [Abstract].
63. Vojtek, A.B., and J.A. Cooper. 1995. Rho family members: activators of MAP kinase cascades. *Cell.* 82: 527-529 [Medline].
64. Xie, H., and S. Clarke. 1994. Protein phosphatase 2A is reversibly modified by methyl esterification of its C-terminal leucine residue in bovine brain. *J. Biol. Chem.* 269: 1981-1984 [Abstract].

This article has been cited by other articles:

- Pawlak, M. R., Scherer, C. A., Chen, J., Roshon, M. J., Ruley, H. E. (2000). Arginine N-Methyltransferase 1 Is Required for Early Postimplantation Mouse Development, but Cells Deficient in the Enzyme Are Viable. *Mol. Cell. Biol.* 20: 4859-4869 [Abstract] [Full Text]

- Abstract of this Article
- **Reprint (PDF) Version of this Article**
- Similar articles found in:
  - JCB Online
  - PubMed
- PubMed Citation
- This Article has been cited by:
- Search Medline for articles by:
  - Cimato, T. R. || Aletta, J. M.
- Alert me when:
  - new articles cite this article
- Download to Citation Manager

Home | Help | Feedback | Subscriptions | Archive | Search | Table of Contents

Neuron, Vol. 24, 659–672, November, 1999

Cypin: A Cytosolic Regulator of PSD-95 Postsynaptic Targeting

Bonnie L. Firestein,* Jay E. Brenman,*
Chiye Aoki,† Ana M. Sanchez-Perez,*
Alaa El-Din El-Husseini,* and David S. Bredt*‡
*Department of Physiology
University of California, San Francisco
San Francisco, California 94143
†Center for Neural Science
New York University
New York, New York 10003

Summary

Postsynaptic density 95 (PSD-95/SAP-90) is a membrane associated guanylate kinase (GK) PDZ protein that scaffolds glutamate receptors and associated signaling networks at excitatory synapses. Affinity chromatography identifies cypin as a major PSD-95-binding protein in brain extracts. Cypin is homologous to a family of hydrolytic bacterial enzymes and shares some similarity with collapsin response mediator protein (CRMP), a cytoplasmic mediator of semaphorin III signalling. Cypin is discretely expressed in neurons and is polarized to basal membranes in intestinal epithelial cells. Overexpression of cypin in hippocampal neurons specifically perturbs postsynaptic trafficking of PSD-95 and SAP-102, an effect not produced by overexpression of other PDZ ligands. In fact, PSD-95 can induce postsynaptic clustering of an otherwise diffusely localized K^+ channel, Kv1.4. By regulating postsynaptic protein sorting, cypin may influence synaptic development and plasticity.

Introduction

Synaptic transmission requires spatial and functional assembly of neurotransmitter receptors and associated signal transduction machinery at synaptic sites. Although mechanisms for the assembly of postsynaptic protein networks remain largely unknown, recent studies demonstrate an important structural role for postsynaptic density 95 (PSD-95 or SAP-90) and related membrane-associated guanylate kinase (MAGUK) proteins (Kim et al., 1995; Komau et al., 1995; Brenman et al., 1996a). PSD-95 was originally isolated as a major protein constituent of the PSD (Cho et al., 1992; Kistner et al., 1993). Molecular cloning indicated that PSD-95 is highly homologous to two previously identified junctional proteins, mammalian zonula occludens 1 (ZO-1) and *Drosophila* discs-large (DLG), that occur at epithelial cell tight junctions (Woods and Bryant, 1991; Willott et al., 1993). In addition to their roles at sites of epithelial cell contact, MAGUKs also mediate important physiological functions at the synapse, as demonstrated by genetic analyses. Mutations of DLG, which is expressed at the fly glutamatergic neuromuscular junction, disrupt postsynaptic structure and result in abnormally large synaptic currents (Lahey et al., 1994; Budnik et al., 1996). Targeted disruption of PSD-95 in mice does not grossly alter synaptic structure but does perturb hippocampal synaptic plasticity and, presumably through this mechanism, blocks spatial learning (Migaud et al., 1998).

MAGUK proteins regulate synaptic function by acting as molecular scaffolds to organize and cluster signaling machinery at the synapse. PSD-95 and the related neuronal MAGUKs, PSD-93 (or Chapsyn-110), SAP-97 (or hDLG), and SAP-102, all contain multiple protein–protein interaction motifs, including three N-terminal PDZ domains, an SH3 domain, and a C-terminal region homologous to yeast guanylate kinase (GK) (Muller et al., 1995, 1996; Brenman et al., 1996a; Kim et al., 1996). PDZ domains from PSD-95 bind to N-methyl-D-aspartate-(NMDA-) type glutamate receptors (Komau et al., 1995), Shaker (Kim et al., 1995), and inward rectifier K^+ channels (Cohen et al., 1996), and to certain peripheral membrane proteins containing a C-terminal PDZ-binding consensus sequence (tSXV), Thr/Ser-X-Val/Ile-COOH (Komau et al., 1997; Sheng and Wyszynski, 1997). PDZ domains also bind to internal motifs in target proteins, as exemplified by the PDZ–PDZ domain interaction between PSD-95 and neuronal nitric oxide synthase (Brenman et al., 1996a; Hillier et al., 1999). By aggregating neurotransmitter receptors and downstream enzymes, PSD-95 and other multi-PDZ proteins act as molecular scaffolds that mediate subcellular protein compartmentalization to ensure the selective activation of different signal transduction cascades within a single cell (Craven and Bredt, 1998; Scott and Zuker, 1998).

To help understand the assembly of synaptic signaling complexes, recent studies have evaluated mechanisms for the postsynaptic clustering of PSD-95. These analyses show that postsynaptic targeting requires not only PDZ domains but also N-terminal palmitoylation of PSD-95, the extended SH3 region, and a short C-terminal targeting motif (Arnold and Clapham, 1999; Craven et al., 1999). Some important implications of this work are that (1) assembly of synaptic complexes is an active process that requires multiple levels of protein sorting and (2) postsynaptic targeting cannot simply be explained by PDZ domain–target interactions.

Here, we sought to identify proteins that regulate intermediate steps in postsynaptic protein sorting by isolating detergent-soluble PSD-95-interacting proteins. Using a PSD-95 protein column for affinity chromatography, we have isolated a highly abundant novel protein, cypin (cytoplasmic PSD-95 interactor), which is a PSD-95 binding protein in brain extracts. Cypin terminates with a PDZ-interacting sequence, −SSSV*, that has a unique binding requirement for PDZ domains 1 and 2 (PDZ1 and PDZ2). Through this interacting domain, cypin associates with multiple members of the PSD-95 family. Cypin is expressed throughout the body but occurs at the highest levels in neurons and intestinal epithelial cells, in which cypin protein is polarized to the

‡To whom correspondence should be addressed (e-mail: bredt@itsa.ucsf.edu).

EXHIBIT C

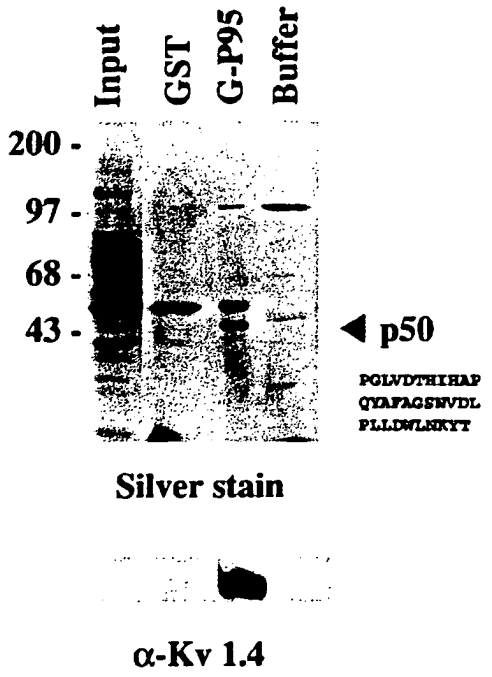


Figure 1. Purification of Cypin

(Top) PSD-95-GST coupled to Affigel-10 beads was used for affinity chromatography of solubilized brain extract. Brain extract loaded onto a GST-Affigel-10 column and a PSD-95-GST-Affigel-10 column loaded with buffer alone served as controls. Silver staining shows a single band of 50 kDa that is specifically retained by the PSD-95 beads. A 30 amino acid peptide sequenced from this band is shown. (Bottom) Western blotting shows that the K^+ channel Kv1.4 also specifically binds to the PSD-95 column.

basal membrane. Ultrastructural studies in brain indicate that cypin occurs in dendritic shafts and in the neck of synaptic spines but is typically excluded from the spine head and PSD. Consistent with this cytosolic distribution, overexpression of cypin in hippocampal neurons disrupts the synaptic clustering of PSD-95 and SAP-102. This is a specific effect of cypin, as overexpression of other diffusely expressed PDZ ligands does not affect the synaptic targeting of PSD-95. On the contrary, PSD-95 can induce postsynaptic clustering of the cotransfected K^+ channel Kv1.4.

Results

Purification of Cypin, a PSD-95-Binding Protein in Brain

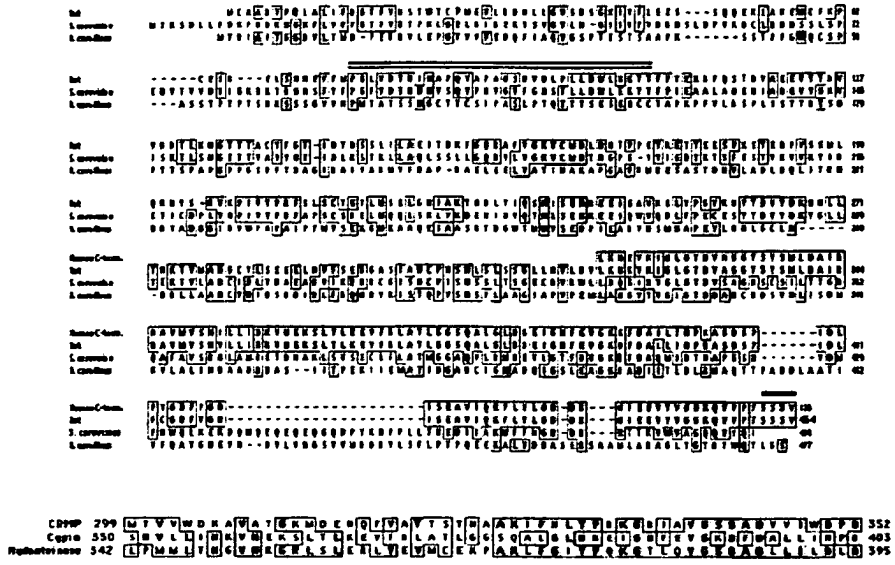
To identify PSD-95-binding proteins in brain, we generated a PSD-95–GST (glutathione S-transferase) fusion protein for affinity chromatography. We applied detergent-solubilized rat brain extracts to the PSD-95 column and in parallel to a control GST column. Both columns were washed extensively, and bound proteins were eluted with 4 M urea and were analyzed by SDS–PAGE. Silver staining revealed a single band of ~50 kDa that was retained by the PSD-95 column but not by the control GST column (Figure 1, arrow). Tubulin, the most abundant protein in the brain extract, bound nonspecifically to both columns (Figure 1A). A small amount of the PSD-95 fusion protein eluted from the column and also eluted from a control PSD-95 column loaded with solubilization buffer alone.

Previous studies have identified a number of PSD-95-binding proteins, and we expected these should also be present in the eluate fractions. Western blotting indicated that known PSD-95-binding proteins, including neuronal nitric oxide synthase (nNOS), Kv1.4, and NMDA receptor 2B, were all purified by the PSD-95 column (Figure 1; data not shown). However, these proteins are apparently of much lower abundance in Triton-soluble extracts than is the 50 kDa protein, and they were not apparent by silver staining.

To determine the structure of this PSD-95-binding protein, hereafter named cypin, we pooled the appropriate column fractions and purified the 50 kDa protein band by preparative SDS–PAGE. Following electrophoresis, proteins were transferred to a polyvinylidene difluoride (PVDF) membrane, and the 50 kDa band was excised. The band was digested with cyanogen bromide, and the released peptides were treated with o-phthalaldehyde, which blocks all peptides that lack an N-terminal proline. Microsequencing following this blocking procedure yielded a single unequivocal peptide sequence of 30 amino acids (Figure 1). The sequence of this peptide perfectly matched the conceptual translation of an expressed sequenced tag (EST) in the public database. This EST was used to probe a rat brain cDNA library, and overlapping clones encoding a novel protein of 50 kDa were obtained (Figure 2A). The overall sequence of cypin shares ~40% identity with an evolutionarily conserved gene family of proteins of unknown function (Figure 2A). Members of this family have been identified as open reading frames in the genomes of organisms from several kingdoms, including prokaryotes, fungi, and archea bacteria. Cypin also shares homology with the family of atrazine hydrolases, bacterial enzymes that dechlorinate the herbicide atrazine (Shao et al., 1995). An EST encoding the C-terminal 136 amino acids of human cypin shows 94.1% sequence identity with the rat cDNA (Figure 2A). In addition, Anthrobacter D-hydantoinase, an atrazine hydrolase, and cypin share homology with collapsin response mediator protein (CRMP) (Figure 2A).

Northern blotting with a cypin cDNA probe identifies a single band of 6.4 kb in mRNA purified from rat brain (Figure 2B). In addition, we detect cypin message in rat liver and spleen as well as in human brain, placenta, liver, and kidney (Figure 2B). Cypin mRNA is not present in human heart, lung, skeletal muscle, or pancreas (Figure 2B).

We developed a polyclonal cypin antiserum, which recognizes a 50 kDa band in crude tissue extracts (Figure 2C). The antiserum faintly cross-reacts with nonspecific bands of 80 and 110 kDa in skeletal muscle; 60, 68, and 95 kDa in brain; and 120 kDa in heart (Figure 2C). The distribution of the 50 kDa cypin protein band is generally similar to the pattern of cypin mRNA expression (Figures 2B and 2C). The highest levels of cypin protein are detected in extracts from kidney, liver, lung, brain, and spleen (Figure 2C). Lower levels of cypin are present in extracts from placenta, heart, and skeletal muscle, and no cypin was expressed in pancreas. It is curious that high levels of cypin protein are found in rat lung, whereas cypin mRNA is not found in human lung. We do not yet understand the root of this difference.



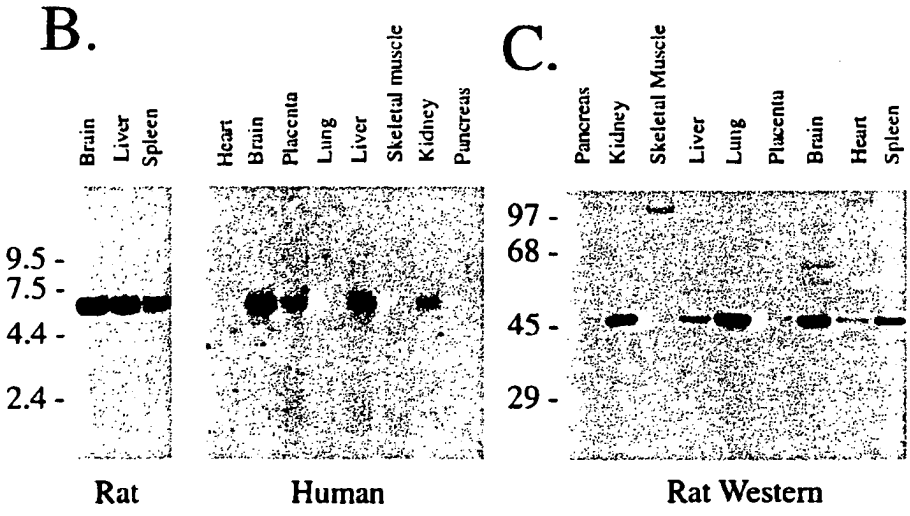


Figure 2. cDNA Cloning of Cypin and Characterization of Cypin Expression

([A], Top) Alignment of rat cypin with a homologous protein from *Saccharomyces cerevisiae*, atrazine hydrolase from *Rhodococcus corallinus* with the conceptual translation of an EST encoding human cypin.

([A], Bottom) Sequence alignment shows a region of significant homology between CRMP, cypin, and Arthrobacter D-hydantoinase.

(B) mRNA for cypin is widely expressed in rat and human tissues. Northern blotting reveals that cypin mRNA (~6.4 kb) occurs in rat brain, liver, and spleen. In mRNA purified from human tissues, cypin is expressed in brain, placenta, liver, and kidney but is not detected in heart, lung, skeletal muscle, or pancreas.

(C) Western blotting shows that cypin protein is expressed at high levels in kidney, liver, lung, brain, and spleen. Cypin is expressed at lower levels in skeletal muscle, placenta, and heart and is absent from pancreas.

Cypin Interacts with the First and Second PDZ Domains of PSD-95 via a C-terminal PDZ Binding-Site

As cypin is an abundant PSD-95 binding protein in detergent-soluble brain extracts, we sought to characterize its mode of interaction with PSD-95. GST fusion protein affinity chromatography was performed using the PDZ1, PDZ2, PDZ1 and PDZ2, PDZ1–PDZ3, SH3, and GK domains, as well as full-length PSD-95. Figure 3A shows that cypin binds to PDZ1 and PDZ2 together but not to either PDZ domain alone. Binding to PDZ1 and PDZ2 is quantitatively similar to cypin binding to full-length PSD-95 (Figure 3B). In addition, cypin does not bind to the SH3 or GK domains (Figure 3B; data not shown). As

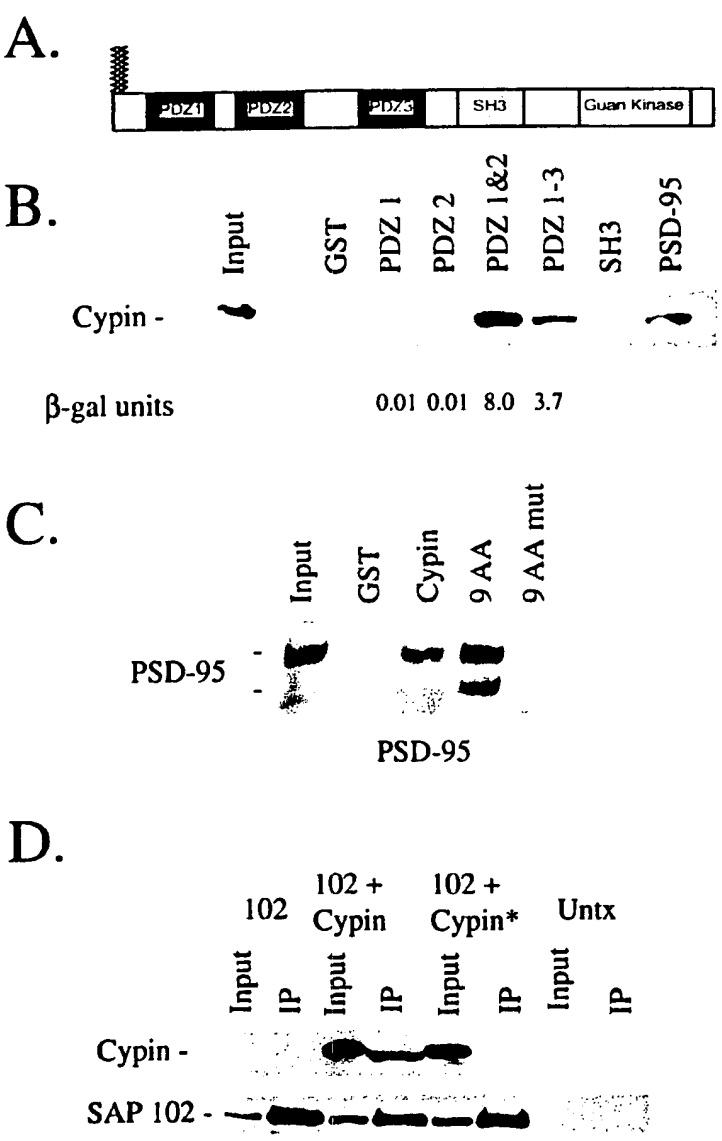


Figure 3. Cypin Interacts with PDZ1 and PDZ2 of PSD-95 through a C-Terminal −SSSV* Consensus Sequence

(A) Schematic domain map of PSD-95.

(B) GST fusion protein affinity chromatography shows that cypin binding to PSD-95 requires PDZ1 and PDZ2. Cypin does not bind to either PDZ domain individually or to the SH3 domain from PSD-95. Input represents 10% of the brain extract used for the pull-down assay. Retained cypin protein was detected by immunoblotting. Yeast two-hybrid analysis confirms that cypin binds to PDZ1 and PDZ2 of PSD-95 but not PDZ 1 or 2 alone. Protein interaction was measured and quantified by colorimetric assay of β-galactosidase enzymatic activity, and values for the various constructs are shown beneath the blot.

(C) Cypin binds to PSD-95 through a C-terminal consensus sequence. GST fusion protein affinity chromatography shows that PSD-95 binds to the last nine amino acids of cypin (cypin and 9 AA) but not to GST. Mutation of the critical serine (−2) and valine (0) to alanine blocks PSD-95 binding to cypin (9 AA mut).

(D) Deletion of the last four amino acids of cypin attenuates association with SAP-102 in vivo. COS7 cells were transfected with constructs encoding the indicated proteins. SAP-102 protein was immunoprecipitated, and association of cypin or cypin mutant (Cypin*) lacking the last four amino acids was determined by immunoblotting.

further support for cypin binding to both the PDZ1 and PDZ2 of PSD-95, we cotransformed yeast with expression vectors encoding various PDZ domains from PSD-95 together with full-length cypin (Figure 3B). Complementing the affinity chromatography data, cypin binds to PDZ1 and PDZ2, whereas cypin fails to interact with either PDZ domain alone.

Conspicuous in the cypin sequence is a C-terminal consensus (−SSSV*), which conforms to the canonical tSXV motif for binding to PSD-95 PDZ domains (Kim et al., 1995; Kornau et al., 1995). To determine whether this C-terminal sequence is necessary and sufficient for binding to PSD-95, we performed affinity chromatography using fusion proteins of full-length cypin or the last nine amino acids of cypin to evaluate binding to PSD-95 from rat brain (Figure 3C). We found that a GST fusion protein of cypin selectively retains PSD-95 and that a GST fusion protein containing only the final nine amino acids of cypin is sufficient to bind an equivalent amount of PSD-95 (Figure 3C). In contrast, a construct of the C-terminal nine amino acids in which the critical serine (−2 position) and terminal valine were mutated to alanine does not bind to PSD-95 (Figure 3C). We also determined if mutation of the last four amino acids of cypin disrupts association with MAGUK proteins in vivo. We found that SAP-102 coimmunoprecipitates with wild-type but not mutant cypin (Figure 3D). Reciprocally, immunoprecipitation of cypin, but not cypin lacking the final four amino acids, coimmunoprecipitated PSD-95 and SAP-102 (data not shown).

To test further the specificity of the PSD-95–cypin interaction, we performed a reverse yeast two-hybrid screen of 7×10^6 colonies from a rat brain cDNA library using cypin as bait. The screen yielded two independent clones; remarkably, both cypin-interacting clones encoded MAGUK proteins. The first clone encoded the

three PDZ domains of PSD-95 (amino acids 17–427); the second encoded PDZ1 and PDZ2 of SAP-102 (amino acids 63–376). These data confirm the strong and specific interaction between cypin and MAGUK family members.

Cypin Cofractionates with Members of the PSD-95 Family of MAGUKs

As all members of the PSD-95 family can bind the C-terminal −SSSV* consensus sequence, we sought to determine which MAGUKs might interact with cypin in brain. The subcellular fractionation of cypin in brain homogenates was first compared with that of various MAGUK proteins. Cypin is concentrated in synaptosomes and is further enriched in the synaptic vesicle fraction (Figure 4A). As previously described, PSD-95 is enriched in the heavy synaptic membranes. A lower band corresponding to a C-terminal fragment of PSD-95 (Topinka and Bredt, 1998) is enriched in the synaptic vesicle fraction (Figure 4A). SAP-97 is found in the synaptic membrane fraction and is highly enriched in the cytosolic fraction. PSD-93 fractionates in a manner similar to that of PSD-95 and is most enriched in the heavy synaptic membrane fraction. Finally, SAP-102 has a similar profile to that of cypin; SAP-102 is also enriched in the synaptic membrane fraction (Figure 4A). The efficiency of fractionation was verified with synaptophysin, a synaptic vesicle marker that is enriched in the synaptic vesicle fraction (Figure 4A). These results show that cypin does fractionate with synaptic plasma membranes and vesicles and potentially could be present both in post- and presynaptic fractions, together with PSD-95/-93 and SAP-97/-102, respectively.

Cypin Associates with MAGUK Proteins in Brain

Association of cypin with members of the PSD family was assessed more directly by immunoprecipitation. Rat brain membranes were solubilized with 0.5% deoxycholate, and protein complexes were immunoprecipitated with two independent antisera to PSD-95. Both sera precipitate cypin together with PSD-95 (Figure 4B, Immune 1, Immune 2). Quantitation of the immunoprecipitates showed that 30% membrane-bound cypin is associated with PSD-95. As a control, we found that no precipitation of PSD-95 or cypin occurred when we used a preimmune serum (Figure 4B, Preimmune 1). Boiling of brain extracts in SDS prior to immunoprecipitation disrupted coimmunoprecipitation of cypin with PSD-95 (Figure 4B, Immune 1 + SDS). In a reciprocal set of experiments, we immunoprecipitated cypin from solubilized membranes from rat brain and probed for coprecipitation of PSD-95 family members. We found that neuronal MAGUK proteins PSD-95, SAP-97, and SAP-102 coimmunoprecipitate with cypin (Figure 4C). As a control, we found that calcium/calmodulin kinase II (CaMKII), an abundant postsynaptic protein, does not coprecipitate with cypin (Figure 4C).

Cypin Is Enriched in Discrete Neuronal Populations in the Rat Brain

To identify neuronal pathways that might be regulated by cypin, we mapped the cellular distribution of cypin in brain. Cypin protein is abundant in extracts from cerebral cortex, hippocampus, olfactory bulb, and corpus striatum but is absent from cerebellum, spinal cord, and medulla (Figure 5A). In situ hybridization demonstrated that cypin mRNA is selectively enriched in a subset of neurons in regions of the brain that express high cypin protein levels (Figure 5B). The highest levels of cypin mRNA are found in rat cerebral cortex and hippocampus and in a stripe of cells in the superior colliculus. Lower levels of cypin are detected diffusely in the corpus striatum (Figure 5B). Like cypin protein, cypin mRNA is absent from cerebellum, medulla, and brainstem (Figure 5B).

Immunohistochemical staining also reveals a neuronal distribution for cypin protein (Figure 5B). Diffuse staining of neuropil is apparent in the cerebral cortex, striatum, and hippocampus. The highest levels of cypin protein are found in a stripe of neurons in the superior colliculus and in interneurons scattered throughout the hippocampus (Figure 5Ca and 5Cb). Preabsorption of the antibody with antigen (20 μg/ml) blocks all staining (Figures 5Cc and 5Cd). These cell populations in the superior colliculus and hippocampus are intensely stained in their cell bodies and throughout their cell processes. Cypin-positive neurons scattered in the hippocampus extensively colocalize with nNOS (Bredt et al., 1991), indicating that they are inhibitory interneurons (Figures 5Ce and 5Cf).

Cypin Protein Is Polarized to the Basal Membrane of Intestinal Epithelial Cells

Although selectively enriched in neurons in brain, cypin also occurs in certain peripheral tissues. In situ hybridization and immunohistochemical staining of numerous peripheral tissues were used to survey which nonneuronal cells express cypin (data not shown). This analysis revealed that the highest levels of cypin throughout the body occur in the epithelial cells of intestine. In situ hybridization demonstrates the discrete expression of cypin message in epithelial cells in small intestine (Figures 5Da and 5Dc). Cypin mRNA expression is highest in differentiated epithelial cells at the tips of the intestinal villi, whereas expression is lower in immature cells in the crypts, and cypin is absent from submucosal and muscular layers. Hybridization with a probe encoding cypin sense sequence yielded no staining (Figures 5Db and 5Dd). Immunohistochemical staining demonstrates that cypin protein expression in intestinal epithelial cells is polarized and occurs exclusively at the basal membrane (Figures 5De and 5Df).

Cypin Localizes to Both Pre- and Postsynaptic Sites in Cortical Neurons

To assess the subcellular localization of cypin in neurons, we performed immunoelectron microscopy on sections of rat cerebral cortex. Examination of peroxidase-labeled sections revealed that the highest densities of cypin occur discretely, in a subset of cortical neuronal processes, with only rare labeling of astrocytic elements. Within neurons, immunoreactivity is found most commonly within the cytoplasm and at the plasma membrane of dendritic shafts. Though PSDs are occasionally coated with the peroxidase label, cypin immunoreactivity often occurs only in the dendritic shaft and

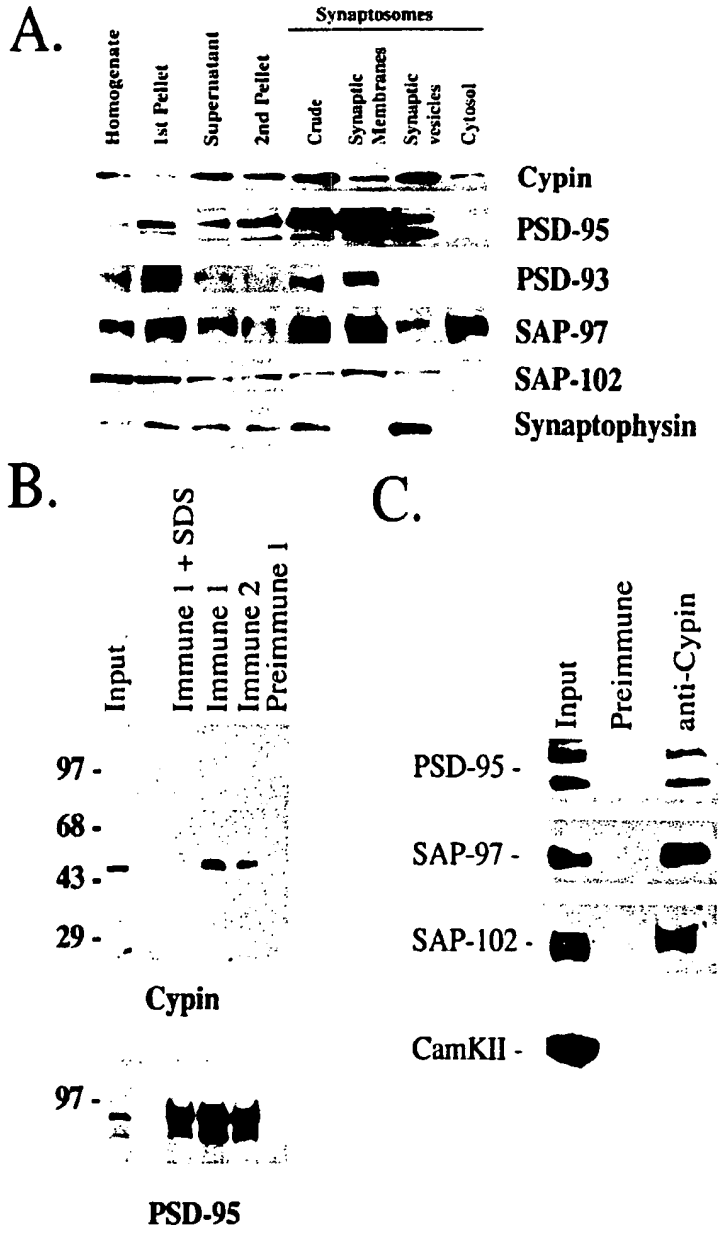


Figure 4. Cypin Cofractionates and Associates with MAGUK Proteins in Brain

(A) Subcellular fractionation of brain was performed as described in the Experimental Procedures. Cypin is concentrated in synaptosomes, synaptic plasma membranes, and the synaptic vesicle fraction, which is enriched in synaptophysin. Like cypin, PSD-95 is enriched in synaptosomes and in synaptic plasma membranes, and a lower band corresponding to a C-terminal fragment of PSD-95 is enriched in the synaptic vesicle fraction. Other MAGUK proteins, PSD-93, SAP-97, and SAP-102, are also enriched in the synaptosomal fractions.

(B) Cypin specifically coimmunoprecipitates with PSD-95 from brain. Rat brain membranes were solubilized with 0.5% deoxycholate and were immunoprecipitated with two independent antisera against PSD-95. Both sera avidly precipitated cypin, together with PSD-95 (Immune 1 and Immune 2). Preimmune serum (Preimmune 1) does not coimmunoprecipitate cypin. Boiling the brain extract in SDS (Immune 1 + SDS) disrupts coimmunoprecipitation of cypin with PSD-95.

(C) PSD-95 and other MAGUK proteins also occur in cypin immunoprecipitate. Brain homogenate solubilized with 1% Triton X-100 was immunoprecipitated with a cypin antiserum. Immunoblotting shows that PSD-95, SAP-97, and SAP-102 coimmunoprecipitate with cypin, whereas CaMKII does not.

neck of the spine and is typically excluded from the spine head and PSD (Figure 6A). At shaft synapses, cypin immunoreactivity is most concentrated upon perisynaptic saccules that aggregate adjacent to the weakly labeled postsynaptic membrane (Figure 6B). In addition to labeling dendritic shafts, cypin also is concentrated in a subset of axonal processes. Within axons, cypin is concentrated at the synaptic terminals, and the reaction product is often seen coating synaptic vesicles (Figure 6C). This is consistent with subcellular fractionation, which shows that cypin is expressed both pre- and postsynaptically.

Because the peroxidase label diffuses, we also used silver-intensified preembedding immunogold, to allow more precise subcellular localization of cypin. Again, we found that cypin is most prevalent within the cytoplasm of dendritic shafts. Immunogold particles rarely occur at the PSD. Instead, the immunoreactivity is more often found along the plasma membrane adjacent to the PSD and upon cytoplasmic perisynaptic saccules (Figures 6D and 6E).

Cypin Regulates Postsynaptic Sorting of MAGUK Proteins

As cypin is enriched in dendritic shafts and perisynaptic domains, we asked whether cypin might regulate the postsynaptic sorting of associated MAGUK proteins. To address this, we evaluated the postsynaptic trafficking of PSD-95 and SAP-102 in hippocampal neurons. Hippocampi were dissected from late gestation embryonic

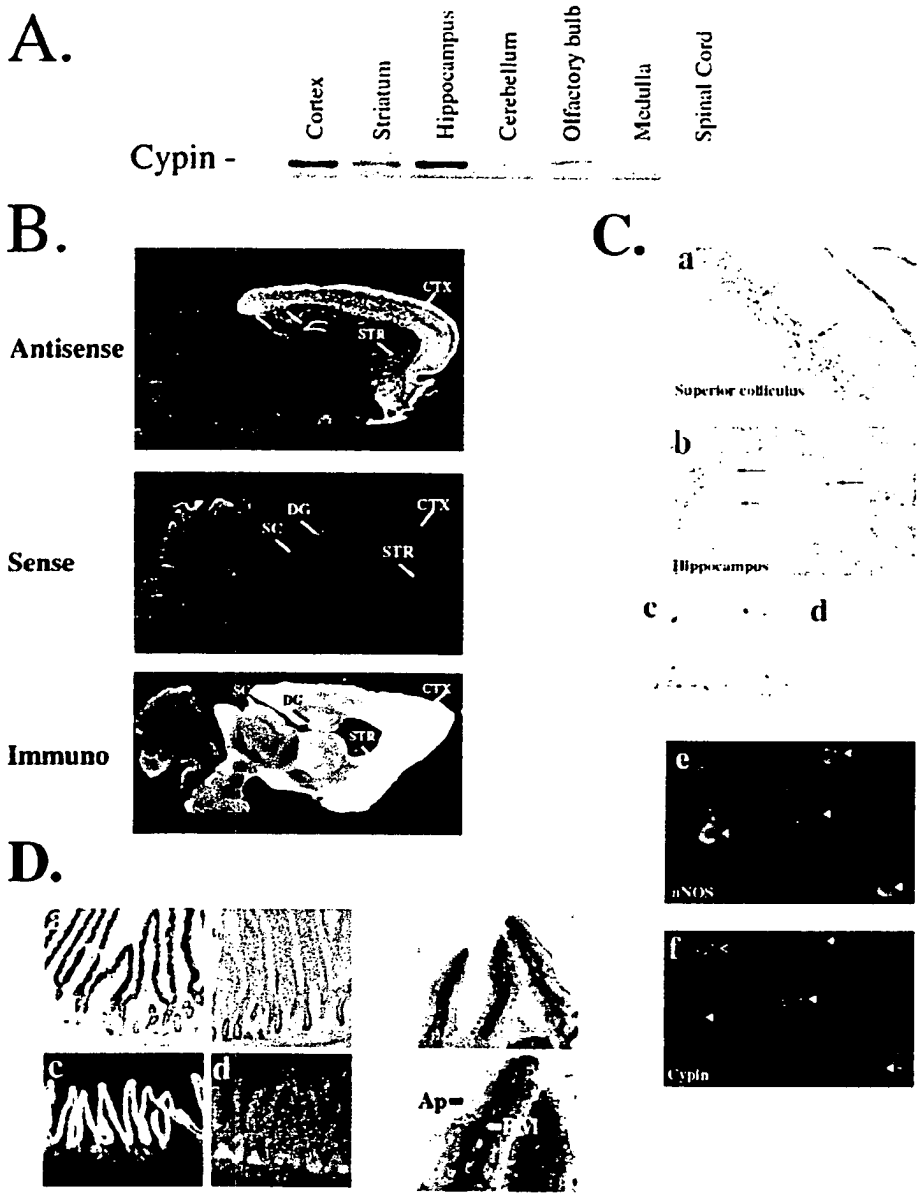


Figure 5. Cypin Is Selectively Expressed in a Subpopulation of Forebrain Neurons and Occurs at the Basal Membrane of Intestinal Epithelial Cells

(A) Western blotting shows that cypin protein is expressed in cortex, hippocampus, striatum, and olfactory bulb but not in cerebellum, medulla, or spinal cord.

(B) In situ hybridization of a sagittal rat brain section with an antisense probe shows high expression of cypin mRNA in cerebral cortex (CTX) and dentate gyrus (DG) and lower expression in the corpus striatum (STR). No specific signal is detected with a sense probe. Immunohistochemical staining of a sagittal brain section reveals a similar pattern of protein expression in discrete neuronal populations of forebrain and no cypin protein expression in cerebellum.

(Ca) Cypin is enriched in cell soma and processes of a layer of neurons in superior colliculus.

(Cb–Cd) Immunohistochemistry reveals that cypin is highly expressed in hippocampal neurons, and preabsorption of the cypin antibody with antigen (20 μg/ml) blocks staining (Cd).

(Ce and Cf) In hippocampus, cypin is highly expressed in interneurons, which mostly colocalize with nNOS (closed arrowheads). A small percentage of cypin-positive neurons in hippocampus are nNOS negative (open arrowhead, [Cd]).

(D) Cypin expression is restricted to the basal membrane of polarized intestinal epithelial cells. An antisense probe to cypin hybridizes to the epithelial cells (Ep) of the intestine (bright field, [Da], dark field, [Dc]), whereas the submucosa (Sub) lacks cypin mRNA. No labeling is detected with a sense probe (bright field, [Db]; dark field, [Dd]). Immunohistochemistry shows cypin protein is polarized to the basal membrane (BM) and is excluded from the apical domain (Ap) of intestinal epithelial cells (200×, [De]; 400×, [Df]).

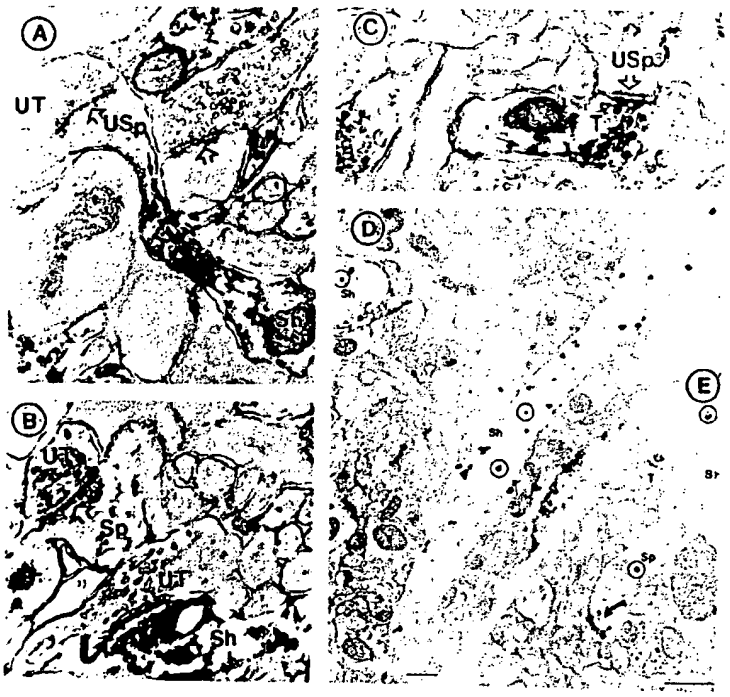


Figure 6. Cypin Immunoreactivity at Synaptic and Nonsynaptic Portions of Neuronal Processes

Electron micrographic immunohistochemical staining of cerebral cortex visualized and labeled with either diaminobenzidine (A–C) or silver-intensified gold (SIG) (D and E).

(A) The open arrow points to an unlabeled postsynaptic density within a spine head (Usp) that is connected to an immunoreactive spine neck (between the two small arrows) emanating from an immunoreactive dendritic shaft (Sh).

(B) Cypin immunoreactivity occurs diffusely distributed within a dendritic shaft (Sh). The curved arrow points to the postsynaptic membrane that is adjacent to an intensely immunoreactive smooth endoplasmic reticulum. The small arrow points to immunoreactivity along the nonsynaptic portion of the plasma membrane.

(C) Cypin immunoreactivity occurs along the nonsynaptic portion of the plasma membrane (small arrow), surrounding vesicles, and along the presynaptic membrane of an axon terminal, T. The open arrow points to the unlabeled postsynaptic density of a spine (Usp).

(D) SIG labeling away (circles) and along (arrowheads) the plasma membrane of two dendritic shafts.

(E) A dendritic spine (Sp) emanating from a dendritic shaft exhibits immunoreactivity along the postsynaptic membrane adjacent to the postsynaptic density (arrowhead) and immediately beneath the postsynaptic density (large curved arrow). As seen in dendritic shafts, immunoreactivity occurs away from and along the plasma membrane. Open arrow points to a postsynaptic density devoid of cypin immunoreactivity.

Abbreviation: T, unlabeled terminals. Scale bar, 1 μm (A–C); and 500 nm (D and E).

rats, neurons were dissociated and transfected in suspension, and cultures were maintained for 14 days in vitro. In mature cultures, PSD-95 and SAP-102 are targeted to dendrites and are clustered at discrete spots Figure 7. These clusters likely reflect postsynaptic sites, as they colocalize with synaptophysin (Craven et al., 1999; data not shown). Overexpression of cypin, which shows a diffuse expression pattern in the dendrites of transfected cells, reduces the number of PSD-95 clusters (Figure 7). To quantitate this effect, we counted the number of synaptic clusters per micrometer of dendrite length (Table 1). We found that transfection with cypin causes a 2-fold reduction in the density of PSD-95 synaptic clusters, from 1.8 clusters per micrometer to 0.9 clusters per micrometer (Table 1). The effect of cypin on SAP-102 targeting is even more striking. Overexpression of cypin completely abolishes synaptic clustering of SAP-102, from 1.8 clusters per micrometer to 0 clusters per micrometer (Table 1). Rather than being clustered at the synapse, SAP-102 becomes diffusely distributed in the dendrites of the transfected neurons (Figure 7). As a control, we found that expression of a cypin mutant lacking the last four amino acids does not affect clustering of PSD-95 (1.8 clusters per micrometer to 1.9 clusters per micrometer) or SAP-102 (1.8 clusters per micrometer to 1.8 clusters per micrometer) (Figure 7; Table 1). We considered the possibility that cypin may influence synaptic clustering of PSD-95 and SAP-102 by altering the total number of synaptic sites. To assess this, we stained transfected cultures with a presynaptic marker, synaptophysin, and found that transfection with cypin does not alter the total density of synapses (data not shown). Thus, cypin appears to specifically regulate the synaptic targeting of associated MAGUK proteins.

To determine whether overexpression of any soluble protein containing a PDZ-binding motif might disrupt MAGUK clustering, we transfected hippocampal cultures with a cDNA encoding cyan fluorescent protein fused to the PDZ-binding motif EKLSSIESDV (CFP–tSXV). Although CFP–tSXV binds to the PDZ2 of PSD-95 in vitro, as expected (data not shown), it has no effect on the clustering of PSD-95 or SAP-102 (Figure 8B; Table 1).

Finally, we asked whether a diffusely expressed transmembrane MAGUK-binding partner might disrupt the clustering of PSD-95 or SAP-102. For this, hippocampal cultures were transfected with a cDNA encoding the K^+ channel Kv1.4. When transfected alone, Kv1.4 occurs diffusely throughout the neuronal soma and in both the axonal and dendritic processes (Figure 8A). However, when cotransfected with either PSD-95 or SAP-102, Kv1.4 clusters at postsynaptic sites (Figure 8A). In the cotransfections, PSD-95 itself retains its normal expression at somatodendritic clusters and is largely excluded from axons. To determine whether the postsynaptic clustering of Kv1.4 targeting requires the PDZ-binding tSXV motif, we designed a Kv1.4 construct lacking the C-terminal intracellular region (Kv1.4 trunc). As seen in Figure 8A, PSD-95 and SAP-102 do not target this truncated Kv1.4 to postsynaptic sites.

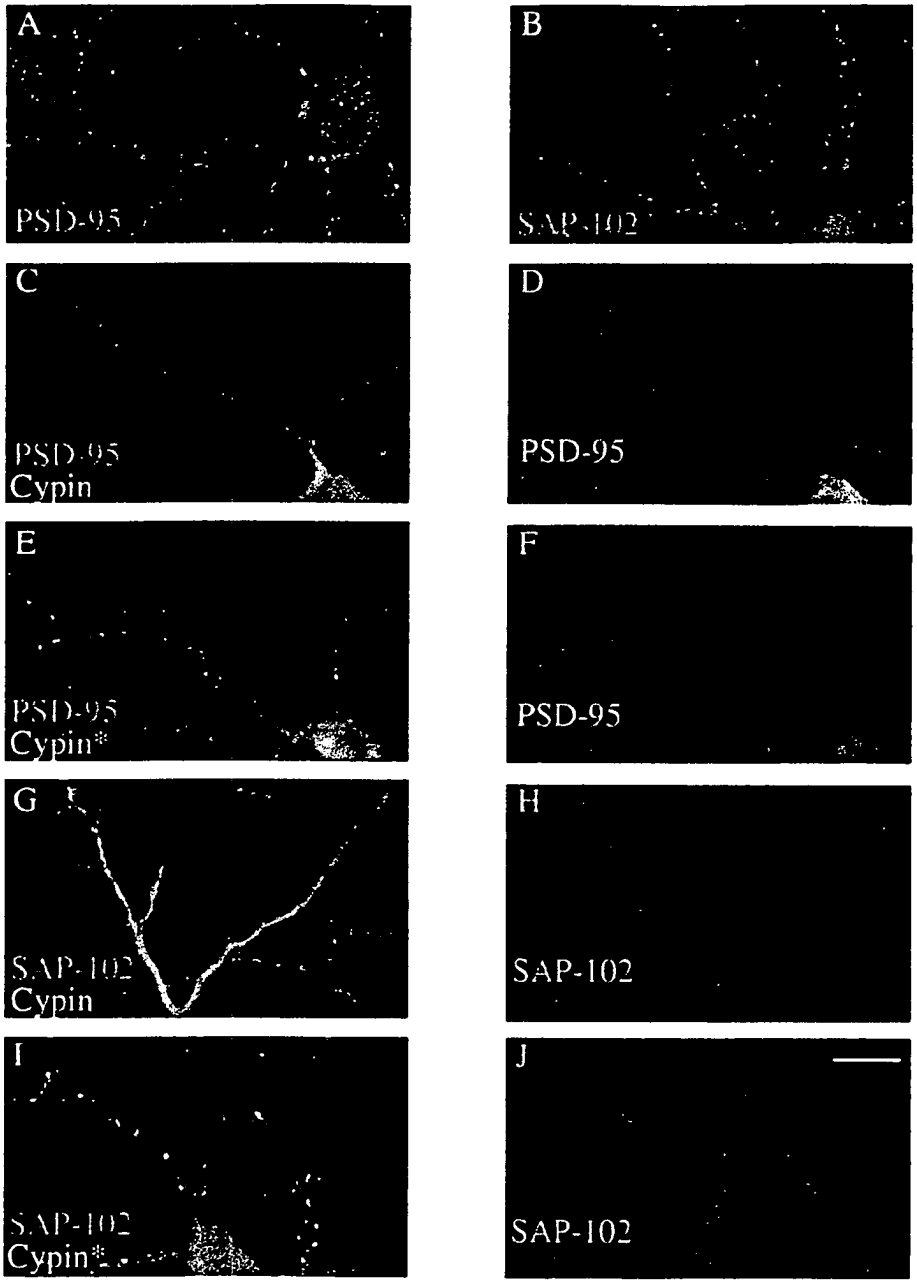


Figure 7. Cypin Regulates Postsynaptic Clustering of PSD-95 and SAP-102 in Primary Hippocampal Neurons

Following 11–14 days in vitro, PSD-95 and SAP-102 occur clustered at synaptic sites (green, [A and B]). Transfection with cypin (red, [C, D, G, and H]) reduces the number of PSD-95 clusters (green, [C]) and eliminates 102 clusters (green, [G]). Transfection with a cypin mutant (Cypin*) lacking the last four amino acids (red, [E, F, I, and J]) has no effect on the clustering of PSD-95 (green, [E]) or SAP-102 (green, [I]). Scale bar, 10 μm.

Discussion

Cypin as a Soluble PSD-95-Binding Protein in Brain

Our affinity chromatography procedure isolated cypin as the most abundant protein in detergent-soluble brain extracts that can bind to PSD-95. The C terminus of cypin Ser-Ser-Ser-Val* binds to PSD-95 with an affinity similar to that of other tSXV ligands (data not shown). Therefore, the unique isolation of cypin by our procedure suggests that cypin is more abundant in brain extracts than other PSD-95 ligands. What are the implications of such a bountiful PDZ ligand in brain? One explanation is that cypin occurs in numerous subcellular compartments in neurons, whereas most other PDZ ligands occur only at synaptic sites (Sheng and Wyszynski, 1997; Kennedy, 1998; O'Brien et al., 1998; Craven et al., 1999).

Table 1. Dendritic Clustering in Cotransfected Neurons

	PSD-95	SAP-102
Alone	1.8 ± 0.1	1.8 ± 0.1
+ Cypin	0.9 ± 0.05*	0 ± 0**
+ Cypin*	2.0 ± 0.2	1.9 ± 0.1
+ PDZ-binding CFP	1.9 ± 0.06	1.9 ± 0.1

The number of clusters per micrometer of dendrite is presented. Cypin* is devoid of the last C-terminal amino acids. CFP–tSXV is CFP fused to the PDZ-binding motif EKLSSIESDV. Four measurements per cell were averaged for four cells per group. Values are mean ± SEM values. Asterisk, $p < 0.01$ compared with PSD-95 alone, and double asterisk, $p < 0.001$ compared with SAP-102 alone (ANOVA, with Bonferroni correction for multiple comparison).

In addition to its enrichment at perisynaptic dendritic domains, cypin also occurs in axons and is concentrated in a subset of nerve terminals (Figure 6). In these different neuronal domains, cypin has the opportunity to interact with distinct MAGUK proteins. Indeed, we find a robust association of cypin with SAP-97 (Figure 4), which is concentrated in axons and nerve terminals. This high abundance of cypin is also consistent with a structural role. Most previously identified PDZ domain ligands for PSD-95 are ion channels and enzymes (Kim et al., 1995; Kornau et al., 1995; Brenman et al., 1996b). These types of signaling molecules have primarily catalytic functions and are typically expressed at relatively low levels. Cypin, on the other hand, appears to mediate certain actions by interaction with MAGUK proteins and would therefore need to occur at high levels in neurons.

Cypin Is Widely Distributed and Highly Conserved in Phylogeny

Proteins similar to cypin are encoded in the genomes of organisms from several kingdoms, including prokaryotes, fungi, and archea bacteria. Although these gene products highly related to cypin have no identified function, cypin also shares significant homology (27.5% identity over 131 amino acids) with the family of atrazine chlorohydrolases (AtzA) present in *Pseudomonas*, *Rhodococcus*, and certain other bacteria. In fact, *cypin* is more similar to these atrazine-catabolizing enzymes than is any other mammalian gene. Atrazine is a chlorinated triazine herbicide used for controlling grassy weeds. Several microorganisms can metabolize atrazine, and the first step involves dechlorination by AtzA.

Why might this abundant PSD-95-binding protein in brain share homology with an ancient family of heterocycle-metabolizing enzymes? It may be relevant that PSD-95 itself contains a domain highly homologous to GK. The purine ring of guanosine is structurally similar to triazine, and cypin may cooperate with PSD-95 in guanine metabolism. It is alternatively possible that cypin metabolizes heterocyclic neurotransmitters derived from histidine, tryptophan, or adenine. It is interesting to note that AtzA and cypin share homology with CRMP (Goshima et al., 1995) and with the *Caenorhabditis elegans* uncoordinated 33 (*unc-33*) gene (Li et al., 1992), which themselves are similar to the bacterial enzyme D-hydantoinase (Figure 2). CRMP mediates intracellular responses to semaphorin III and is required for semaphorin III–induced growth cone retraction (Goshima et al., 1995). These actions of CRMP are presumably mediated by hydrolysis of neuronal hydantoins or pyrimidines (Wang and Strittmatter, 1997). As cypin and CRMP both may hydrolyze nitrogen-based heterocycles, they may interact to regulate neuronal development. Finally, very recent studies demonstrate that the glycine receptor–interacting protein gephyrin plays both a structural role in regulating glycine receptor clustering and a metabolic role in synthesizing the molybdoenzyme cofactor (Feng et al., 1998). Cypin may also be a dual function protein, with both structural and catabolic activities in neurons.

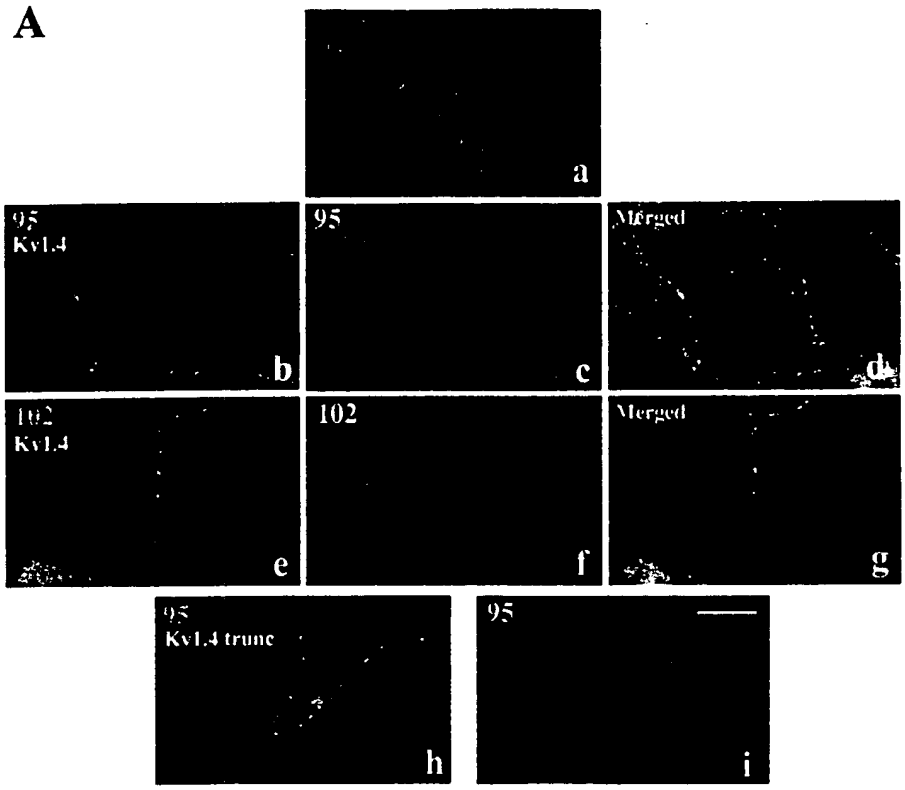
Cypin Localizes to Specific Subcellular Domains in Neuronal and Nonneuronal Cells

Outside the brain, the highest levels of cypin occur in intestinal epithelial cells. Within epithelial cells of both the large and small intestines, cypin localizes to the basal region of the lateral plasma membrane, and cypin is excluded from the apical membrane region. This basolateral membrane distribution is similar to that of SAP-97, one of the identified binding partners for cypin (Muller et al., 1995; Reuver and Garner, 1998; Wu et al., 1998). Efficient lateral membrane targeting of SAP-97 requires PDZ1 and PDZ2 (Wu et al., 1998), which also compose the site for cypin binding. Cypin may therefore regulate SAP-97 localization and function in epithelial cells.

In brain, cypin occurs discretely in several neuronal populations in forebrain, though cypin is completely absent from hindbrain and spinal cord. In several forebrain regions, including the hippocampus, cypin is expressed at the highest levels in inhibitory interneurons. This finding fits with emerging evidence that the PSD-95-NMDA receptor complex is differentially regulated in excitatory and inhibitory neurons. Whereas PSD-95 and NMDA receptor subunits themselves occur ubiquitously at excitatory synapses in both glutamatergic and GABAergic neurons, several PSD-95-associated proteins are selectively expressed in specific cell types. In hippocampal neurons, synGAP is highly enriched in excitatory neurons (Zhang et al., 1999), whereas several other MAGUK-binding proteins, including GKAP (Rao et al., 1998), citron (Zhang et al., 1999), and nNOS (Bredt et al., 1991) resemble cypin and are enriched in inhibitory neurons. This alternative composition of the PSD can begin to explain the differences in signal transduction downstream of NMDA receptors observed in GABAergic versus glutamatergic neurons. For example, NMDA receptor–dependent long-term depression (LTP) does not occur in hippocampal interneurons (Maccaferri and McBain, 1996). As cypin is a highly abundant protein that presumably can compete with NMDA receptors and other synaptic ligands for interaction with PSD-95, cypin may regulate plasticity in hippocampal interneurons by altering PSD-95-associated signal transduction cascades activated by glutamate.

Cypin as a Regulator of Postsynaptic Clustering of MAGUKs

Functional expression in neurons indicates that cypin can regulate synaptic clustering of MAGUK proteins. This is particularly relevant in light of recent studies indicating that postsynaptic targeting of PSD-95 is an



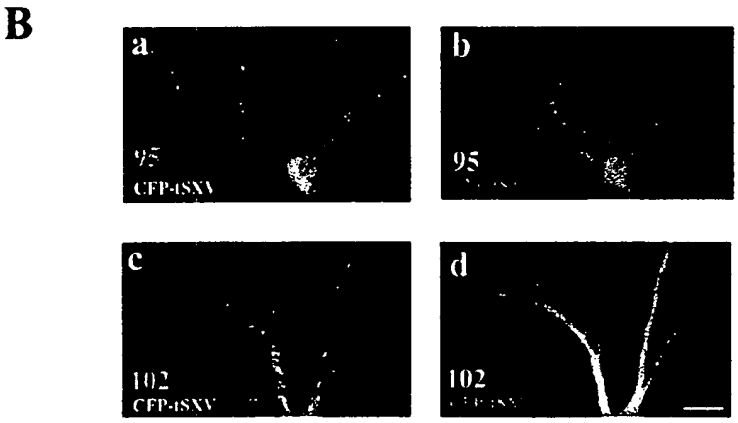


Figure 8. PSD-95 and SAP-102 Can Mediate Synaptic Clustering of Kv1.4

(A) Kv1.4 is diffusely expressed thoughout the neuron when transfected alone (red, [Aa]). Coexpression with PSD-95 (green, [Ab–Ad]) or SAP-102 (green, e–g) targets Kv1.4 to postsynaptic sites (red, [Ac and Af]). Kv1.4 lacking the C-terminal intracellular region (Kv1.4 trunc; red, [Ai]) is not targeted to postsynaptic sites by PSD-95 (green, [Ah]).

(B) Cyan fluorescent protein fused to the PDZ-binding motif EKLSSIESDV (CFP–tSXV; blue, [Bb and Bd]) is diffusely expressed and does not affect postsynaptic targeting of PSD-95 (green, [Ba]) or SAP-102 (green, [Bd]). Scale bar, 10 μm.

active cellular process that relies on N-terminal palmitoylation, PDZ domains, an extended SH3 region, and a C-terminal targeting motif (Arnold and Clapham, 1999; Craven et al., 1999). Whereas the palmitoylation and the C-terminal motif are likely involved in cellular sorting pathways (A. E. E. and D. S. B., unpublished data; S. E. Craven and D. S. B., unpublished data), the PDZ domain interactions seem more likely to anchor PSD-95 at the PSD, which is replete with PDZ ligands. Consistent with this notion, overexpression of cypin, a nonsynaptic PDZ ligand in hippocampal cultures, disrupts synaptic targeting of PSD-95 and SAP-102, and this process relies on the PDZ-binding site in cypin. This negative regulation of MAGUK protein clustering is a unique function

of cypin and is not reproduced by merely overexpressing a decoy PDZ-binding ligand. This suggests that specific structural features of cypin mediate the observed effects on postsynaptic clustering; the putative catalytic region of cypin that shares homology with atrazine hydrolase and CRMP is a possible candidate. Future studies on systematically mutagenizing cypin should help address this issue.

Finally, we find that overexpression of PSD-95 or SAP-102 in cultured neurons recruits Kv1.4 to postsynaptic sites, which represents the first formal demonstration of postsynaptic clustering by MAGUK proteins in neurons. This result must be compared with data in a very recent study (Arnold and Clapham, 1999), in which coexpression of Kv1.4 and PSD-95 did not yield postsynaptic ion channel clustering. Instead, these investigators found that Kv1.4 caused axonal targeting of PSD-95. Our experiments differ from those of Arnold and Clapham in several respects, including the neurons studied (hippocampal versus cortical), the methods for neuronal culturing (dispersed cultures versus slice cultures), and the times for protein expression after transfection (12 days versus 2 days). What is clear from these two studies is that the subtle interplay between PSD-95 and its PDZ-binding partners determines the cellular distribution of the complex. As such, cypin is as yet the first gene shown to negatively regulate synaptic targeting of this complex.

It is important to note that hippocampal cultures consist primarily of excitatory and therefore cypin-depleted neurons, such that transfection of cypin into these cells markedly changes its cellular level and can explain the dramatic effect on MAGUK clustering. As cypin is an abundant protein expressed in specific neuronal populations, where it may regulate the number of synapses containing PSD-95 protein, the effect we find with transfection into relatively cypin-depleted hippocampal pyramidal cells is likely to reflect its physiological role in other neuronal subpopulations. Postsynaptic protein sorting is central to synaptogenesis, and dynamic recruitment of existing and newly synthesized proteins mediates aspects of LTP. By regulating postsynaptic targeting and interactions with PSD-95, cypin may have important roles in neuronal development and plasticity.

Experimental Procedures

Antibodies

The following primary antibodies were used: mouse monoclonal to nNOS (Transduction Laboratories), rabbit polyclonal to Kv1.4 (generously provided by Dr. Lily Jan, University of California, San Francisco), rabbit polyclonal to PSD-95, guinea pig polyclonal to SAP-97 (Topinka and Bredt, 1998), monoclonal to PSD-95 (Affinity Bioreagents), and monoclonal to green fluorescent protein (Quantum). Immunizing guinea pigs with a GST-SAP-102 fusion protein encoding the first 97 amino acids of SAP-102 raised the polyclonal antiserum to SAP-102. Immunizing rabbits with a full-length GST–cypin fusion protein generated the polyclonal antiserum to cypin. All antisera were affinity purified on columns containing the immunizing antigen linked to Affigel-10 resin.

Protein Purification

Affinity chromatography was performed as previously described (Firestein and Bredt, 1999). Five mg of GST, PSD-95-GST fusion protein, or bovine serum albumin (BSA) (Brenman et al., 1998) were cross-linked per milliliter of Affigel-10 (Biorad). Adult rat brains (25) were homogenized in 10 volumes of ice-cold buffer B, Triton X-100 was added to 1%, and proteins were extracted for 1 hr at 4°C. The homogenate was spun at 12,000 × g for 10 min, and incubating with 2 ml of BSA-Affigel-10 beads precleared the supernatant. The BSA beads were removed, and 100 ml of extract was incubated with 0.5 ml of PSD-95-Affigel-10 at 4°C for 3 hr. As controls, 10 ml of extract was incubated with 50 μl of GST-Affigel-10, and 10 ml of buffer B containing 1% Triton X-100 was incubated with 50 μl of PSD-95-Affigel-10. Extract/beads were then loaded into the appropriately sized columns and were washed with 100 column volumes of buffer B containing 0.5% NP40, 1% Triton X-100, and 1 mM phenylmethylsulfonyl fluoride (PMSF) followed by buffer B containing 1% Triton X-100. Proteins were eluted consecutively with 5 column volumes of buffer B containing 1% Triton X-100 and 2.5 M urea (twice); 5 column volumes of buffer D containing 1% Triton X-100 and 4 M urea; and finally, 2.5 ml of buffer B containing 1% SDS. Eluate fractions were precipitated with 15% trichloroacetic acid, proteins were resolved by 8% SDS–PAGE, and bands were visualized with silver.

The 2.5 M urea and 4 M urea fractions were combined, and proteins were separated on an 8% polyacrylamide gel and transferred to PVDF membrane (Immobilon P, Millipore). The membrane was stained with Coomassie blue to visualize bands, the desired protein bands were excised, and ProSeq (Salem, MA) performed protein sequencing, which yielded a single peptide sequence of 30 unambiguous amino acids.

GST fusion protein "pull-down" assays (Brenman et al., 1996b) and synaptosomal fractionation (Firestein and Bredt, 1999) were done as previously described.

cDNA Cloning and mRNA Analysis

An EST was labeled and used to screen 10^6 phage plaques from a rat brain cDNA library (Stratagene). Five clones were isolated and sequenced on both stands. These clones contained a single large open reading frame, which predicted a protein of 454 amino acids that contained the 30 amino acid peptide sequenced from cypin protein.

For Northern blotting, rat RNAs were isolated by the guanidine isothiocyanate/CsCl method, and mRNA was selected, using oligo dT sepharose. mMRNA was separated on formaldehyde agarose gels and transferred to nylon membranes. Equal amounts of mRNA were loaded as determined by ethidium bromide staining. The human Northern blot was from Clontech. Filters were hybridized with a random primed ^{32}P probe generated with cypin cDNA as template. After hybridization, membranes were washed at high stringency—68°C, 0.1% SSC, 0.1% SDS—and exposed to X-ray film at −70°C.

In situ hybridization used ^{35}S-labeled RNA probes and was performed exactly as described (Brenman et al., 1996b). Tissue sections were exposed to X-ray film for 4 days and then were dipped into photographic emulsion (Kodak NTB-2) and exposed for 14 days.

Primary Neuronal Culture and Transfection

Neuronal cultures were prepared from hippocampi of rat embryos at 18 days gestation. The hippocampi were dissociated with papain followed by brief mechanical trituration. Cells were plated on poly-D-lysine-treated glass coverslips (12 mm in diameter) at a density of 200–250 cells/mm^2. Cultures were plated and maintained in neurobasal media supplemented with B27, penicillin, streptomycin, and L-glutamine. Hippocampal cultures were transfected by a lipid-mediated gene transfer as previously described (Craven et al., 1999). Human embryonic kidney (HEK) and COS7 cells were transfected with LipofectAMINE Plus (GIBCO-BRL).

Immunohistochemistry

Light microscopic immunohistochemistry was done as previously described (Brenman et al., 1996a). Affinity-purified anti-cypin antiserum was used at 1 μg/ml, and preabsorption with antigen (20 μg/ml) abolished all staining.

For electron microscopy, rats were perfused with 4% paraformaldehyde and 3% acrolein 0.1 M phosphate buffer (pH 7.4). Sodium borohydride was used to terminate tissue fixation. Brain sections (40 μm) containing primary visual cortex were blocked in phosphate buffer containing 1% BSA. Following an overnight incubation with

affinity-purified anti-cypin antibody (1 μg/ml), sections were processed for immunolabeling with the avidin–biotin–HRP complex with the ABC Elite kit (Vector) and 3,3′-diaminobenzidine (Aldrich Chemicals) and hydrogen peroxide as substrates. Alternatively, sections were processed for preembedding colloidal gold immunolabeling, using 1 nm gold-conjugated anti-rabbit Immunoglobulin G as the label. A silver intensification kit (Amersham) was then used to enlarge the 1 nm gold particles to >5 nm for light and electron microscopic detection (Chan et al., 1990). These sections were postfixed with 1% glutaraldehyde for 10 min, then with 1% osmium tetroxide for 1 hr, and stained en bloc with 4% uranyl acetate. Sections were dehydrated for infiltration with the EMBED 812 resin (EM Sciences) for ultrathin sectioning. Sections (80 nm) were examined without lead citrate counterstaining.

Immunoprecipitation

Frozen brains were homogenized in 10 volumes of TEE (50 mM Tris–HCl [pH 7.4], 150 mM NaCl, and 1 mM PMSF). Insoluble proteins were collected by centrifugation at 20,000 × g for 15 min, and the pellet was extracted in TEE with 0.5% deoxycholate for 1 hr at 4°C. After preclearing with protein A–Sepharose for 1 hr, lysates were immunoprecipitated with 10 μg/ml polyclonal anti-cypin antibody or preimmune antiserum linked to protein A–Sepharose (Pharmacia) for 2 hr at 4°C. Beads were washed with TEE containing 1% Triton X-100, and bound proteins were eluted in SDS loading buffer and analyzed by immunoblotting. For some experiments, the brain pellet was extracted with 0.2% SDS; this solubilized extract was heated to 95°C for 5 min, and Triton X-100 was then added to 1% prior to immunoprecipitation.

Transfected HEK293 or COS7 cells were lysed in 1 ml of TEE with 1% Triton X-100 for 30 min at 4°C. Lysates were cleared of debris and used for coimmunoprecipitation with 10 μl/ml of polyclonal guinea pig anti-SAP-102 or monoclonal anti-PSD-95 (045, Affinity Bioreagents) as described above for brain tissue.

Yeast Two-Hybrid System

Full-length cypin cDNA was subcloned into pGTB9 to generate a GAL4-binding domain fusion. This construct was cotransformed into yeast strain HF7c with a library of rat brain cDNAs fused to the GAL4 activation domain (Clontech). The transformation mixture was plated onto a synthetic dextrose plate lacking tryptophan, leucine, and histidine. Growth was monitored during a 5 day incubation at 30°C, and color was measured by a β-galactosidase colorimetric filter assay. Interacting clones were rescued, retransformed to confirm interaction, and sequenced. Complementary cDNAs encoding PDZ domains from PSD-95 were cloned into pGAD-10 to generate GAL4 activation domain fusion constructs. Yeast (strain SY256) was cotransformed with cypin, and the various PSD-95 deletion constructs and protein interactions were quantified by colorimetric β-galactosidase assay.

Acknowledgments

The authors thank Joshua F. Kilbridge, Alice Elste, Mary Chan, and Joel Rosen for excellent technical assistance; Sarah E. Craven for insightful discussions; and Elizabeth Bellochio for advice on synaptosomal fractionation. This work was supported by a National Institute for Child Health and Human Development National Research Service Award (B. L. F.); a Spinal Cord Research Foundation Postdoctoral Fellowship (B. L. F); an American Heart Association Predoctoral Fellowship (J. E. B.); a Human Frontier Science Program Award (C. A.); National Institutes of Health grants R01-EY08055 and R01-NS36017 (C. A. and D. S. B.); National Science Foundation grants RCD 92-53750 and 94015013 (C. A. and D. S. B.); and by additional grants from the New York University Research Challenge Fund (C. A.), the EJLB (D. S. B.), and the Culpeper and Beckman Foundations (D. S. B.).

Received July 9, 1999; revised September 20, 1999.

References

Arnold, D.B., and Clapham, D.E. (1999). Molecular determinants for subcellular localization of PSD-95 with an interacting K^+ channel. Neuron *23*, 149–157.

Bredt, D.S., Glatt, C.E., Hwang, P.M., Fotuhi, M., Dawson, T.M., and Snyder, S.H. (1991). Nitric oxide synthase protein and mRNA are discretely localized in neuronal populations of the mammalian CNS together with NADPH diaphorase. Neuron *7*, 615–624.

Brenman, J.E., Chao, D.S., Gee, S.H., McGee, A.W., Craven, S.E., Santillano, D.R., Huang, F., Xia, H., Peters, M.F., Froehner, S.C., and Bredt, D.S. (1996a). Interaction of nitric oxide synthase with the postsynaptic density protein PSD-95 and α-1 syntrophin mediated by PDZ motifs. Cell *84*, 757–767.

Brenman, J.E., Christopherson, K.S., Craven, S.E., McGee, A.W., and Bredt, D.S. (1996b). Cloning and characterization of postsynaptic density 93 (PSD-93), a nitric oxide synthase interacting protein. J. Neurosci. *16*, 7407–7415.

Brenman, J.E., Topinka, J.R., Cooper, E.C., McGee, A.W., Rosen, J., Milroy, T., Ralston, H.J., and Bredt, D.S. (1998). Localization of postsynaptic density–93 to dendritic microtubules and interaction with microtubule-associated protein 1A. J. Neurosci. *18*, 8805–8813.

Budnik, V., Koh, Y.H., Guan, B., Hartmann, B., Hough, C., Woods, D., and Gorczyca, M. (1996). Regulation of synapse structure and function by the Drosophila tumor suppressor gene *dlg*. Neuron *17*, 627–640.

Chan, J., Aoki, C., and Pickel, V.M. (1990). Optimization of differential immunogold-silver and peroxidase labeling with maintenance of ultrastructure in brain sections before plastic embedding. J. Neurosci. Methods *33*, 113–127.

Cho, K.O., Hunt, C.A., and Kennedy, M.B. (1992). The rat brain postsynaptic density fraction contains a homolog of the Drosophila discs-large tumor suppressor protein. Neuron *9*, 929–942.

Cohen, N.A., Brenman, J.E., Snyder, S.H., and Bredt, D.S. (1996). Binding of the inward rectifier K^+ channel Kir 2.3 to PSD-95 is regulated by protein kinase A phosphorylation. Neuron *17*, 759–767.

Craven, S.E., and Bredt, D.S. (1998). PDZ proteins organize synaptic signaling pathways. Cell *93*, 495–498.

Craven, S.E., Husseini, A.E., and Bredt, D.S. (1999). Synaptic targeting of the postsynaptic density protein PSD-95 mediated by lipid and protein motifs. Neuron *22*, 497–509.

Feng, G., Tintrup, H., Kirsch, J., Nichol, M.C., Kuhse, J., Betz, H., and Sanes, J.R. (1998). Dual requirement for gephyrin in glycine receptor clustering and molybdoenzyme activity. Science *282*, 1321–1324.

Firestein, B.L., and Bredt, D.S. (1999). Interaction of neuronal nitric oxide synthase and phosphofructokinase-M. J. Biol. Chem. *274*, 10545–10550.

Goshima, Y., Nakamura, F., Strittmatter, P., and Strittmatter, S.M. (1995). Collapsin-induced growth cone collapse mediated by an intracellular protein related to UNC-33. Nature *376*, 509–514.

Hillier, B.J., Christopherson, K.S., Prehoda, K.E., Bredt, D.S., and Lim, W.A. (1999). Unexpected modes of PDZ domain scaffolding revealed by structure of nNOS-syntrophin complex. Science, in press.

Kennedy, M.B. (1998). Signal transduction molecules at the glutamatergic postsynaptic membrane. Brain Res. Brain Res. Rev. *26*, 243–257.

Kim, E., Niethammer, M., Rothschild, A., Jan, Y.N., and Sheng, M. (1995). Clustering of Shaker-type K^+ channels by direct interaction with the PSD-95/SAP90 family of membrane-associated guanylate kinases. Nature *378*, 85–88.

Kim, E., Cho, K.-O., Rothschild, A., and Sheng, M. (1996). Heteromultimerization and NMDA receptor clustering activity of chapsyn-110, a novel member of the PSD-95 family of synaptic proteins. Neuron *17*, 103–113.

Kistner, U., Wenzel, B.M., Veh, R.W., Cases-Langhoff, C., Garner, A.M., Appeltauer, U., Voss, B., Gundelfinger, E.D., and Garner, C.C. (1993). SAP90, a rat presynaptic protein related to the product of the Drosophila tumor suppressor gene dlg-A. J. Biol. Chem. *268*, 4580–4583.

Kornau, H.-C., Schenker, L.T., Kennedy, M.B., and Seeburg, P.H. (1995). Domain interaction between NMDA receptor subunits and the postsynaptic density protein PSD-95. Science *269*, 1737–1740.

Kornau, H.-C., Seeburg, P.H., and Kennedy, M.B. (1997). Interaction

of ion channels and receptors with PDZ domains. Curr. Opin. Neurobiol. *7*, 368–373.

Lahey, T., Gorczyca, M., Jia, X.X., and Budnik, V. (1994). The Drosophila tumor suppressor gene *dlg* is required for normal synaptic bouton structure. Neuron *13*, 823–835.

Li, W., Herman, R.K., and Shaw, J.E. (1992). Analysis of the Caenorhabditis elegans axonal guidance and outgrowth gene unc-33. Genetics *132*, 675–689.

Maccaferri, G., and McBain, C.J. (1996). Long-term potentiation in distinct subtypes of hippocampal nonpyramidal neurons. J. Neurosci. *16*, 5334–5343.

Migaud, M., Charlesworth, P., Dempster, M., Webster, L.C., Watabe, A.M., Makhinson, M., He, Y., Ramsay, M.F., Morris, R.G., Morrison, J.H., et al. (1998). Enhanced long-term potentiation and impaired learning in mice with mutant postsynaptic density-95 protein. Nature *396*, 433–439.

Muller, B.M., Kistner, U., Veh, R.W., Cases-Langhoff, C., Becker, B., Gundelfinger, E.D., and Garner, C.C. (1995). Molecular characterization and spatial distribution of SAP97, a novel presynaptic protein homologous to SAP90 and the Drosophila discs-large tumor suppressor protein. J. Neurosci. *15*, 2354–2366.

Muller, B.M., Kistner, U., Kindler, S., Chung, W.K., Kuhlendahl, S., Fenster, S.D., Lau, L.-F., Veh, R.W., Huganir, R.L., Gundelfinger, E.D., and Garner, C.C. (1996). SAP102, a novel postsynaptic protein that interacts with NMDA receptor complexes in vivo. Neuron *17*, 255–265.

O'Brien, R.J., Lau, L.F., and Huganir, R.L. (1998). Molecular mechanisms of glutamate receptor clustering at excitatory synapses. Curr. Opin. Neurobiol. *8*, 364–369.

Rao, A., Kim, E., Sheng, M., and Craig, A.M. (1998). Heterogeneity in the molecular composition of excitatory postsynaptic sites during development of hippocampal neurons in culture. J. Neurosci. *18*, 1217–1229.

Reuver, S.M., and Garner, C.C. (1998). E-cadherin mediated cell adhesion recruits SAP97 into the cortical cytoskeleton. J. Cell Sci. *111*, 1071–1080.

Scott, K., and Zuker, C.S. (1998). Assembly of the Drosophila phototransduction cascade into a signaling complex shapes elementary responses. Nature *395*, 805–808.

Shao, Z.Q., Seffens, W., Mulbry, W., and Behki, R.M. (1995). Cloning and expression of the s-triazine hydrolase gene (trzA) from Rhodococcus corallinus and development of Rhodococcus recombinant strains capable of dealkylating and dechlorinating the herbicide atrazine. J. Bacteriol. *177*, 5748–5755.

Sheng, M., and Wyszynski, M. (1997). Ion channel targeting in neurons. Bioessays *19*, 847–853.

Topinka, J.R., and Bredt, D.S. (1998). N-terminal palmitoylation of PSD-95 regulates association with cell membranes and interaction with K^+ channel Kv1.4. Neuron *20*, 125–134.

Wang, L.H., and Strittmatter, S.M. (1997). Brain CRMP forms heterotetramers similar to liver dihydropyrimidinase. J. Neurochem. *69*, 2261–2269.

Willott, E., Balda, M.S., Fanning, A.S., Jameson, B., Van Itallie, C., and Anderson, J.M. (1993). The tight junction protein ZO-1 is homologous to the Drosophila discs-large tumor suppressor protein of septate junctions. Proc. Natl. Acad. Sci. USA *90*, 7834–7838.

Woods, D.F., and Bryant, P.J. (1991). The discs-large tumor suppressor gene of Drosophila encodes a guanylate kinase homolog localized at septate junctions. Cell *66*, 451–464.

Wu, H., Reuver, S.M., Kuhlendahl, S., Chung, W.J., and Garner, C.C. (1998). Subcellular targeting and cytoskeletal attachment of SAP97 to the epithelial lateral membrane. J. Cell Sci. *111*, 2365–2376.

Zhang, W., Vazquez, L., Apperson, M., and Kennedy, M.B. (1999). Citron binds to PSD-95 at glutamatergic synapses on inhibitory neurons in the hippocampus. J. Neurosci. *19*, 96–108.

GenBank Accession Number

The GenBank accession number for the *cypin* sequence reported in this paper is AA376792.

Note Added in Proof

While this manuscript was in review, a mouse protein highly similar to cypin was purified as a guanine deaminase (J. Biol. Chem. *274*, 8175–8189), suggesting that linked metabolism of guanine nucleotides by cypin and the guanylate kinase domain of PSD-95 may regulate synaptic clustering.

Volume 272, Number 6, Issue of February 7, 1997 pp. 3590-3598

Phosphorylation, Subcellular Localization, and Membrane Orientation of the Alzheimer's Disease-associated Presenilins*

(Received for publication, May 30, 1996, and in revised form, October 1, 1996)

Bart De Strooper [‡§], Monique Beullens [¶], Bart Contreras [‖], Lyne Levesque [—], Katleen Craessaerts [‡], Barbara Cordell [#], Dieder Moechars [‡], Mathieu Bollen [¶], Paul Fraser [**—], Peter St. George-Hyslop [**—] and Fred Van Leuven [‡]**

From the [‡] Experimental Genetics Group and [‖] Molecular Oncology Group, Center for Human Genetics, and the [¶] Laboratory of Biochemistry, Katholieke Universiteit, 3000 Leuven, Belgium, the [**] Center for Research in Neurodegenerative Diseases, Department of Medicine (Neurology) and Medical Biophysics, University of Toronto, Toronto, Ontario M5S 1A8, Canada, and [#] Scios Inc., Sunnyvale, California 94043

ABSTRACT

Presenilins 1 and 2 are unglycosylated proteins with apparent molecular mass of 45 and 50 kDa, respectively, in transfected COS-1 and Chinese hamster ovary cells. They colocalize with proteins from the endoplasmic reticulum and the Golgi apparatus in transfected and untransfected cells. In COS-1 cells low amounts of intact endogeneous presenilin 1 migrating at 45 kDa are detected together with relative larger amounts of presenilin 1 fragments migrating between 18 and 30 kDa. The presenilins have a strong tendency to form aggregates (mass of 100-250 kDa) in SDS-polyacrylamide gel electrophoresis, which can

be partially resolved when denatured by SDS at 37 °C instead of 95 °C. Sulfation, glycosaminoglycan modification, or acylation of the presenilins was not observed, but both proteins are posttranslationally phosphorylated on serine residues. The mutations Ala-246 → Glu or Cys-410 → Tyr that cause Alzheimer's disease do not interfere with the biosynthesis or phosphorylation of presenilin 1. Finally, using low concentrations of digitonin to selectively permeabilize the cell membrane but not the endoplasmic reticulum membrane, it is demonstrated that the two major hydrophilic domains of presenilin 1 are oriented to the cytoplasm. The current investigation documents the posttranslational modifications and subcellular localization of the presenilins and indicates that postulated interactions with amyloid precursor protein metabolism should occur in the early compartments of the biosynthetic pathway.

INTRODUCTION

Alzheimer's disease is a major health problem. Patients suffer from a progressive dementia caused by massive neuronal loss in cortical and hippocampal areas of the brain (1-6). Neuropathological signs of the disease are tangles and amyloid deposits in the brain parenchyma, and amyloid deposits in the brain vasculature. The cause of the sporadic form of the disease is still unknown, although an increased risk is associated with the presence of apolipoprotein allele E4 (6, 7). On the other hand, familial early onset Alzheimer's disease is caused by point mutations in the amyloid precursor protein gene on chromosome 21 (8), in the presenilin 2 (PS2)[1] gene on chromosome 1 (9-11), or, most frequently, in the presenilin 1 (PS1) gene on chromosome 14 (12-15). Amyloid precursor protein (APP) is a type I integral membrane protein and is the precursor of the amyloid peptide, the main component of the senile plaques (1-3). Point mutations in exons 16 and 17 of the APP gene cause alterations in the metabolism of APP. This results in an increased production of intracellular βA4 amyloid peptide containing carboxyl-terminal APP fragments and in an increased secretion of the potentially neurotoxic βA4 peptide (1-3, 16). 63% of the amino acid residues in the sequences of the two presenilins are conserved, which strongly suggests that both proteins are involved in similar functions and have a similar pathogenic role in Alzheimer's disease. Based on computer algorithms, seven membrane spanning domains have been defined (9-15), although the possibility of nine transmembrane domains cannot be ruled out at this moment (17). The amino-terminal domain and the acidic loop domain, located between transmembrane domains six and seven, are hydrophilic and can be alternatively spliced (9, 13). The mutations that cause familial Alzheimer's disease are found all over the protein, but the hydrophilic loop constitutes a "hot spot" with nine different mutations described to date (5, 15). The biological function of the presenilins remains essentially unknown, but, interestingly, the 103 carboxyl-terminal amino acid residues of PS2 can inhibit apoptosis in a "deathtrap" assay (18). Based on the homology with *Caenorhabditis elegans* proteins, roles in intracellular protein sorting and/or intercellular signal transmission have been proposed as well (19, 20).

A hypothetical final common pathway in the pathogenesis of the genetic and sporadic forms of Alzheimer's disease has been postulated, based on the invariable occurrence of the amyloid deposits, the neurofibrillar tangles and the neurodegeneration in all affected brains. The central question is thus how mutations in the presenilin genes can cause this typical neuropathology. Studies indicating an increased production of more amyloidogenic βA4-(1-42) peptide in fibroblasts obtained from patients with presenilin mutations would support the amyloid hypothesis that postulates abnormal βA4 amyloid peptide production and plaque

formation as the pivotal event in the pathogenesis of the disease (2, 21). However, the observed increases in βA4-(1-42) peptide production were relatively small, and the suggested relationship between presenilin mutations and APP metabolism should be further corroborated by experiments showing directly the effect of presenilin mutations on APP processing in transfected cells or in brains of transgenic animals. Moreover, other aspects of APP metabolism such as the production of carboxyl-terminal APP fragments (16) should be investigated in greater detail. The possibility that presenilins interact with the cytoskeleton or exert their effect via apoptotic pathways should not be disregarded at this time (18). In addition to addressing these questions, basic information is needed on the subcellular localization, the posttranslational modifications, and the membrane orientation of the presenilins.

We used COS-1 cells and CHO cells to express PS1 and PS2, as well as Myc-tagged PS1 and PS1-containing mutations that cause familial Alzheimer's disease. The biosynthesis of transfected and untransfected presenilins was studied using immunoblotting or metabolical labeling and immunoprecipitation assays. We demonstrate that transfected presenilins are phosphorylated on serine residues. Using immunofluorescence microscopy, we document the association of presenilin 1 with the early compartments of the biosynthetic pathway and demonstrate the cytoplasmic orientation of the two major hydrophilic domains.

MATERIALS AND METHODS

Constructs

The cDNA coding for mouse PS1 and human PS1, PS2, and PS1 containing Ala-246 → Glu (FAD1) or Cys-410 → Tyr (NIH2) mutations have been described (9, 12). A Myc-tagged PS1 fragment was generated by PCR using primers 5′-CGGGATCCATTATGGAGCAAAAGCTCATTTCTGAAGAGGACTTGACAGAGTTACCTGCACCG-3′ and 5′-GATCACATGCTTGGCGCCATAT-3′ (the sequence coding for the Myc tag is underlined). This fragment was used to replace the *NarI/Bam*HI restriction fragment of PS1 in pSG5. The resulting cDNA codes for PS1 with the Myc tag (EQKLISEEDL) immediately after the initiator methionine, as confirmed by cDNA sequencing. Plasmids containing the cDNA for furin (22, 23) or reticulon/NSP (24, 25) were kindly provided by Dr. J. Creemers, Dr. A. Roebroek, and Dr. W. Van De Ven (Center for Human Genetics, Leuven, Belgium).

Antibodies

Polyclonal rabbit antisera B13, B14, and B15 were raised against peptide p45 (NDNRERQEHNDRRSLC), which is in the amino-terminal domain of PS1 (12). Polyclonal rabbit sera B16 and B17 were raised against peptide p46 (EGDPEAQRRVSKNSKC) situated in the hydrophilic loop domain of PS1. Polyclonal rabbit antiserum 519 was raised against residues KDGQLIYTPFTEDTE(C). This antiserum reacts with a sequence in the second loop domain of PS1 and PS2. The synthetic peptides p45 and p46 were manufactured by Eurogentec (Liège, Belgium), while peptide p519 was synthesized by solid-phase techniques and purified by reverse phase high pressure liquid chromatography in our laboratories.[2] Rabbits were injected every 2 weeks with 100, 200, or 300 μg of peptide in complete Freund's adjuvant, coupled to

keyhole limpet hemocyanin or bovine serum albumin (Pierce) and solubilized in PBS mixed with complete Freund's adjuvant. Monoclonal antibody (mAb) PS1-3, reacting with the peptide RRVSKNSKYNAESTERESQDTVAEN in the hydrophilic loop domain of PS1 mAb 9E10 against the Myc tag (26), was kindly provided by Dr. J. Creemers. mAbs MON160, 161 and 162 and MON148 and 152 against NSP/reticulon and furin have been described (22-25). mAbs against the immunoglobulin-binding protein (BiP) and against the calcium pump Serca 2a (IID8) were purchased from, respectively, StressGen (Victoria, Canada) and Affinity BioReagents (Neshanic Station, NJ).

Cell Culture, Metabolic Labeling, Cell Extraction, and Immunoprecipitation

The fibroblast cell line (designation AG07657, Coriell Institute) from an unaffected individual of the FAD1 lineage was cultured in Dulbecco's modified Eagle's medium supplemented with 20% fetal bovine serum (Life Technologies, Inc.). COS1 cells were cultured as described previously in Dulbecco's modified Eagle's medium/Ham's F-12 with 10% fetal bovine serum (27). CHO cells stably transfected with APP770 were kindly provided by Dr. B. Greenberg (Cephalon, West Chester, PA) and cultured in high glucose Dulbecco's modified Eagle's medium (Life Technologies, Inc.) supplemented with 0.1 mM minimal essential medium (nonessential amino acids), 3% fetal calf serum, and 250 nM methotrexate.

COS cells were transfected using DEAE/dextran (27), while CHO cells were transfected using LipofectAMINE according to the instructions of the manufacturer (Life Technologies, Inc.).

Metabolic labeling was done with 100 μCi of [^{35}S]methionine, 250 μCi of [^{35}S]sulfate, 60 μCi of [^{3}H]glucosamine, 500 μCi of [^{3}H]palmitic acid or [^{3}H]myristic acid, or 500 μCi of [^{32}P]orthophosphate/ml of the appropriate culture medium. All radiolabeled precursors were from Amersham. Cellular labeling was done for 4 h unless otherwise specified, and cell extracts were made as described previously (27). For analysis of phosphorylation, a postnuclear extract was prepared using 0.5% (v/v) Triton X-100 in Tris-buffered saline (TBS) containing proteinase inhibitors (100 units/ml aprotinin, 1 μg/ml pepstatin) and tyrosine and serine/threonine phosphatase inhibitors (1 mM sodium orthovanadate, 5 mM EDTA, 5 mM EGTA, 20 mM NaF). The nuclei were pelleted by centrifugation at 14,000 rpm (15 min) in a cooled Eppendorf centrifuge. Antisera were added to the cell extracts at a 1/250 dilution, and antibody-antigen complexes were collected by incubation with immobilized protein G (Pierce) overnight at 4 °C. The immunoprecipitates were washed using Tris-buffered saline containing 1% (v/v) Triton X-100, 1% (w/v) deoxycholate, and 0.1% (w/v) SDS (27). For analysis of phosphorylation, phosphatase inhibitors as detailed above, were added to all buffers.

Phorbol myristic acid, phorbol dibutyric acid, forskolin, okadaic acid and staurosporin (all from Sigma) were added to the cell cultures during the last 30 min of the metabolic labeling at the indicated final concentrations. For *in vitro* labeling assays, immunoprecipitated PS1 or PS2 bound on protein G beads were incubated with purified protein kinase A or C in Tris-buffered saline, in the presence of 0.1 mM [γ-^{32}P]ATP, 2 mM Mg^{2+} during 60 min at 30 °C. The precipitates were washed and analyzed in SDS-PAGE. Gels were quantitatively analyzed using a PhosphorImager (Molecular Dynamics).

For immunoblotting experiments, cells were scraped in PBS, centrifuged (10 min, 1000 rpm), and solubilized in Laemmli sample buffer. Blots were stained with mAb 9E10 (Myc tag) ascites fluid (1/1000

dilution) or mAb PS1-3 hybridoma supernatant (1/100 dilution) and affinity-purified goat anti-mouse peroxidase-conjugated antibodies (1/10,000; Bio-Rad), using the sensitive ECL system (Amersham).

Phosphoamino Acid Analysis

Immunoprecipitated ^{32}P-labeled PS were size-fractionated in SDS-PAGE, and labeled bands were localized by autoradiography. The proteins were hydrolyzed in 6 N HCl (110 °C, 90 min). The hydrolysate was supplemented with phosphoamino acid standards and analyzed by two-dimensional thin-layer electrophoresis at pH 1.9 and 3.5 in the respective dimensions (28). Phosphoserine, phosphothreonine, and phosphotyrosine were localized using ninhydrin staining, and the radioactive phosphoamino acid residues were visualized by autoradiography.

Immunocytochemistry

Fibroblasts were grown on coverslips coated with mouse collagen IV (Collaborative Biomedical Products; 1 μg/cm^2). Transfected COS-1 or CHO cells were cultured in Lab-TEK chamber slides (Nunc). Cells were washed twice in PBS, fixed in 4% formaldehyde in PBS for 10 min at room temperature, and washed three times in PBS and once in TBS. Cells were permeabilized with 0.02% (v/v) Triton X-100 in TBS for 20 min or with 0.2% (w/v) saponin for 10 min and washed with 0.1% Tween 20 in TBS. Nonspecific binding was blocked with 0.2% cold water fish gelatin, 2% bovine serum albumin, and 2% fetal calf serum (blocking buffer). Cells were probed with affinity-purified primary antibody 519 (1:25) or with immune serum (1:400). Appropriate FITC- and TRITC-conjugated anti-mouse, anti-rat, and anti-rabbit antibodies (Sigma) were used at 1/400 dilution. Preparations were viewed on a Nikon Diaphot 300 or a Zeiss Axiophot UV microscope. Digitized immunofluorescence images were obtained using an LSM419-inverted laser-scanning confocal microscope (Zeiss Inc.) and processed using NIH Image software.

Selective permeabilization of the plasma membrane was obtained by incubating fixed cells in 10 mM Pipes buffer (pH 6.8) containing 0.3 M sucrose, 0.1 M KCl, 2.5 mM $MgCl_2$, 1 mM EDTA, 5 μg/ml digitonin during 15 min at 4 °C (29). Cells were washed with PBS and further processed. A rat monoclonal antibody against the KDEL sequence (30, 31) was used to demonstrate permeabilization of the endoplasmic reticulum membrane.

RESULTS

Immunocytochemical Localization of Presenilin 1 in the Endoplasmic Reticulum and the Golgi Apparatus

The subcellular localization of the presenilins was investigated in permeabilized and fixed fibroblasts using affinity-purified polyclonal antibody 519 raised against the peptide KDGQLIYTPFTEDTE, which is a conserved sequence in the second loop domain of PS1 and PS2 (9, 12). A fine reticular staining in the cytoplasm and a more pronounced perinuclear staining was observed in the cells (Fig. 1*A*), which partially colocalized with the staining obtained with a mAb against BiP, an endoplasmic reticulum marker. Preabsorption of antibody 519 with the peptide antigen resulted in loss of staining, demonstrating the specificity of the observed signals (Fig. 1*C*).

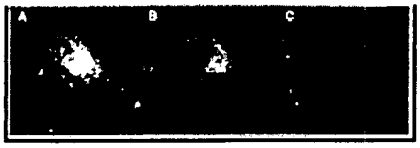


Fig. 1. Immunofluorescent staining of PS1/2 in untransfected fibroblasts. Fibroblasts were fixed with formaldehyde and permeabilized with Triton X-100. Cells were incubated with affinity-purified rabbit antibody 519 against the loop 2 domain of PS1/2 (*panel A*) and mouse mAb against BiP (*panel B*), followed by appropriate TRITC- and FITC-conjugated secondary antibodies. Presenilin staining is observed in the cytoplasm as a reticular pattern (*A*) that partially colocalizes with the staining obtained with the mAb against the endoplasmic reticulum marker BiP (*B*). *Panel C*, fibroblasts stained with antibody 519 preabsorbed with peptide revealed only weak background fluorescence.
[View Larger Version of this Image (32K GIF file)]

The same, lacelike network was seen in COS-1 cells transfected with plasmids containing the cDNA for PS1 and using antiserum B14 against peptide NDNRERQEHNDRRS in the amino-terminal domain of PS1 or antiserum B16 against peptide EGDPEAQRRVSKNSKY in the hydrophilic loop domain of PS1 (Fig. 2, *a* and *c*). Untransfected COS-1 cells remained negative under the experimental conditions used, probably because of the very low levels of endogeneous PS present in these cells (see below). The same pattern of staining was observed with monoclonal antibody IID8 (32) against Serca 2a Ca^{2+}-ATPase (results not shown), and with antibodies against transfected reticulon/NSP (Fig. 2*b*). Both proteins are located in the endoplasmic reticulum and the Golgi apparatus (24). The distribution of PS1 and reticulon/NSP remained identical in double transfected cells (compare *panels a* and *b* in Fig. 2). Transfected APP (Fig. 2*e*) and transfected furin (Fig. 2*d*), in contrast, were mainly found in the Golgi apparatus, as shown previously (33, 34). Similar results were obtained in CHO cells (results not shown).

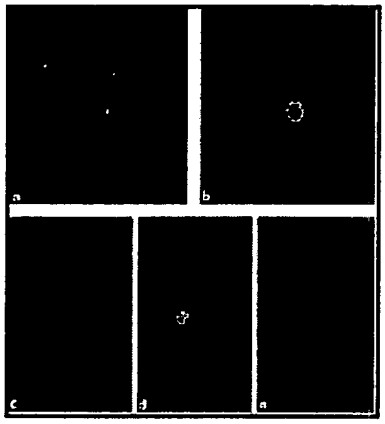


Fig. 2. Immunofluorescent staining of PS1 in transfected COS cells. COS-1 cells were double transfected with plasmids coding for PS1 and reticulon/NSP (*a* and *b*) or single transfected with plasmids coding for PS1, amyloid precursor protein, or furin (*c*, *d*, and *e*). Cells were fixed with formaldehyde and permeabilized with saponin. Immunostaining was done with antiserum B14 against the amino-terminal domain of PS1 (*a*) or B16 against the loop domain of PS1 (*c*), with a mix of monoclonal antibodies MON160-162 against reticulon/NSP (*b*), or MON148 and 152 against furin (*d*), or mAb 22C11 against amyloid precursor protein (*e*), followed by the appropriate TRITC-conjugated (*red*) or FITC-conjugated (*green*) secondary antibodies. Notice the fine reticular pattern obtained with PS1 antibodies in *panels a* and *c*, which distributes with reticulon/NSP (*panel b*) in double transfected cells. *Panel e* demonstrates that amyloid precursor protein is located mainly in the Golgi apparatus, similar to furin (*panel d*).
[View Larger Version of this Image (128K GIF file)]

Biosynthesis of the Presenilins

Metabolic labeling of untransfected or "mock" transfected COS-1 or CHO cells using [^{35}S]methionine, followed by immunoprecipitation of the cell extracts using antibodies 519, B14, or B17 and resolution of the immunoprecipitates in SDS-PAGE, yielded no signals (Fig. 3*A*, *lane 4*) or, after prolonged exposures (2 weeks and more), only nonspecific signals in autoradiography (results not shown). Immunoprecipitation of detergent extracts of COS-1 cells transfected with plasmids containing the cDNA of wild type PS1, in contrast, yielded strong specific signals of radiolabeled protein migrating with an apparent molecular mass of 45 kDa (Fig. 3*A*, *lanes 1-3*). Diffuse protein bands with masses between 100 and 250 kDa were observed to a variable extent (Fig. 3*A*). Similar results were obtained with PS2 cDNA. The main PS2 species migrated, however, slightly more slowly than PS1, resulting in an apparent mass of 50 kDa (see below). Unrelated polyclonal antibodies, or untransfected cells did not yield these bands, while immunoblots of COS-1 cells transfected with Myc-tagged PS1 and stained with the Myc tag-specific mAb 9E10 revealed again the pronounced smearing (Fig. 3*B*). This result clearly demonstrated that the 100-250-kDa protein smears consisted of PS protein, either associated or not associated with other proteins. Essential similar patterns were observed when PS1 FAD1 (Ala-246 → Glu) or PS1 NIH2 (Cys-410 → Tyr) were expressed (Fig. 3*A*). These clinical mutations (12) therefore do not cause major alterations in the biosynthesis of PS1 protein when overexpressed in COS-1 cells (Fig. 3*A*). Independent experiments performed in CHO cells confirmed completely these results (results not shown).

Fig. 3. Biosynthesis of PS1 protein in COS-1 cells. *Panel A*, COS-1 cells were transfected with constructs containing the cDNA coding for wild type PS1 (*PS1 WT*), PS1 FAD1 (Ala-246 → Glu), PS1 NIH2 (Cys-410 → Tyr), or with expression vector alone (*pSG5 control*). Cells were metabolically labeled with [^{35}S]methionine for 4 h. Cells were solubilized, and PS1 was immunoprecipitated with antiserum B17 (1/250). Immunoprecipitates were resolved by 4-20% gradient SDS-PAGE. Molecular size markers are indicated at the *left* in kDa. *Panel B*, Western blotting of COS-1cells transfected with PS1 containing amino-terminally inserted Myc epitope. Cell extracts were electophoresed on a 4-20% gradient acrylamide gel and transferred to a nitrocellulose membrane. mAb 9E10 against the Myc epitope was used to detect PS1. *Panel C*, combined immunoprecipitation and Western blotting of PS1 in untransfected COS cells. Detergent extracts of 10×10^6 untransfected COS-1 cells were made, and PS1 was immunoprecipitated using antiserum B13 (*N-term*) or B17 (*Loop*). In *lanes 1*, *3*, *5*, and *7*, 0.5×10^5 COS-1 cells transfected with wild type PS1 were added to the untransfected cells. Immunoprecipitated material was resolved in 4-20% gradient SDS-PAGE and transferred to a nitrocellulose filter. Filters were reacted with mAb PS1-3 and goat anti-mouse peroxidase-conjugated antibodies (*lanes 1-4*, indicated by *PS1-3*) or with goat anti-mouse peroxidase conjugated antibodies alone (*lanes 5-8*, *−Co*). The mobility of intact PS1 is indicated by an *arrow* at the *right*.
[View Larger Version of this Image (24K GIF file)]

Since the levels of endogeneous PS1 were not detectable using classical immunoprecipitation, a sensitive immunological "sandwich" type assay was developed. PS1 was immunoprecipitated from detergent extracts of 10×10^6 untransfected COS-1 cells using polyclonal PS1 antisera. A small amount of transfected cells (0.5×10^5) was added to the positive control samples. The immunoprecipitates were then resolved in

SDS-PAGE electrophoresis and transferred to a nitrocellulose filter. Immunoprecipitated PS1 was finally detected with mAb PS1-3 raised against peptide RRVSKNSKYNAESTERESQDTVAE in the loop domain of PS1. Weak signals representing intact endogeneous PS1 were revealed in untransfected COS-1 cells (Fig. 3*C*, *lanes 2* and *4*). Endogeneous PS1 migrated with the same mobility (45 kDa) as transfected PS1, as is best seen in the experiments using the amino-terminal domain-specific antiserum B13 (Fig. 3*C*, *lane 2*). Since mAb PS1-3 recognizes an epitope in the carboxyl-terminally located hydrophilic loop, the combination of these antibodies is expected to detect mainly intact PS1 (Fig. 3*C*, *lane 2*). With hydrophilic loop-specific antiserum B17, in contrast, relatively pronounced 18-30-kDa fragments were visualized together with intact PS1. These fragments represent most likely carboxyl-terminal fragments of presenilin (35). Remarkably, the observed fragments were not, or only marginally, increased in COS-1 cells overexpressing PS1 (compare *lanes 3* and *lane 4* in Fig. 3*C*). Pulse-chase experiments on transfected CHO-cells (Fig. 4, *A* and *C*) and COS-1 cells (results not shown) further showed that the transfected 45-kDa PS1 species has a half-life of about 4 h. No fragments were seen at any time point of this assay, either with amino-terminal (B13) or hydrophilic loop-specific (B17) antisera. APP, immunoprecipitated from the same cell extracts (Fig. 4*B*), displayed a much faster turnover (half-life: 2 h), indicating that the relative high expression of PS1 did not interfere with the normal turnover of APP.

Fig. 4. Pulse-chase metabolic labeling of PS1 in transfected CHO cells. CHO cells expressing stably human APP770 were transfected with PS1 and metabolically labeled with [^{35}S]methionine during 10 min in methionine-free medium. Incorporated label was chased by incubation of the cells in complete medium for the indicated time. PS1 was immunoprecipitated from the cell extracts using B13 (*N-term*) or B17 (*Loop*) antiserum and resolved by 4-20% gradient SDS-PAGE (*panel A*). APP was consecutively immunoprecipitated from the same cell extracts using antibody 207 (45, 47) and analyzed in 6% homogeneous SDS-PAGE. Notice the difference in turnover of transfected presenilin 1 and transfected APP. *Panel C* displays a quantitative analysis (mean ± S.E.) of three independent pulse-chase experiments of PS1. Signals were quantitated using a PhosphorImager. The points represent the fraction of PS1 synthesized during a 10-min pulse and remaining after the period of chase as indicated. The results obtained for APP in two experiments are also displayed.
[View Larger Version of this Image (14K GIF file)]

Posttranslational Modification and Aggregation of Presenilins

The problem of the presenilin smears in SDS-PAGE was further investigated. Enzymatic digestion of immunoprecipitated PS1 and PS2 protein with glycosidase F (3 milliunits/μl), *O*-glycanase (60 nanounits/μl), endoglycosidase H (60 nanounits/μl), sialidase (100 nanounits/μl), or combinations of these enzymes did not affect the mobility of the proteins, indicating that no glycosylation of PS1 and PS2 occurred (results not shown). Digestions with heparinase (0.1 milliunits/μl), heparitinase (0.1 milliunits/μl), chondroitinase AC (5 milliunits/μl), and chondroitinase ABC (5 milliunits/μl) also had no effect on the mobility of the protein smears, indicating that PS1 is not modified by glycosaminoglycan chains (Fig. 5). These negative results were independently confirmed by metabolic labeling experiments

using [^{3}H]glucosamine or [^{35}S]O_4. While both radioactive precursors were incorporated in proteins in the cell extracts of the labeled cells, no signal was obtained when transfected PS1 was immunoprecipitated (results not shown). Experiments using [^{3}H]palmitic or [^{3}H]myristic acid demonstrated also that the presenilins did not incorporate fatty acids.

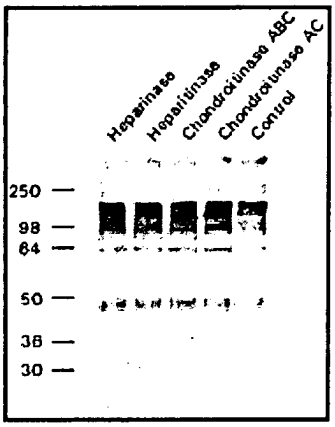


Fig. 5. Presenilins are not modified by glycosaminoglycan chains. PS1 was immunoprecipitated using serum B16 from metabolically labeled COS-1 cells transfected with wild type PS1 and digested with heparinase, heparitinase, chondroitinase ABC, or chondroitinase AC. Digested material was resolved in 12% homogeneous SDS-PAGE. The activity of the heparinase and heparitinase was confirmed using proteoglycans immunopurified from fibroblasts (50). The activity of chondroitinase ABC and AC was demonstrated using a test solution of chondroitinesulfate which was clarified after addition of the enzymes. Molecular size markers are indicated at the *left* in kDa.

[View Larger Version of this Image (39K GIF file)]

Sonication or boiling of the samples in SDS and 6 M urea (data not shown) or extraction of the cells in the presence of 5 mM dithiothreitol did not resolve the aggregates (Fig. 6). Denaturation of the immunoprecipitates at 37 °C instead of at 95 °C reduced the aggregates considerably but not completely (Fig. 6, *lane 2*). This demonstrated that the aggregates consisted mainly, if not exclusively, of PS1 as the radiolabeled species (Fig. 6) and also suggested that the PS aggregates are at least partially produced during the processing of the samples for electrophoresis. It should be noticed, however, that even when freshly prepared material was used and heating of the samples was avoided, protein smears in the 100-250-kDa region remained visible in SDS-PAGE (Fig. 6).

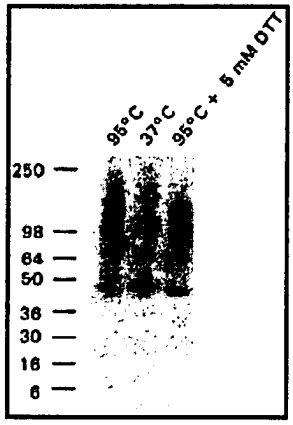


Fig. 6. The presenilin aggregates are temperature-sensitive. Immunoprecipitated PS1 (antiserum B16) from transfected COS cells was denatured for 10 min with SDS and β-mercapthoethanol at 95 °C (*lanes 1* and *3*) or at 37 °C (*lane 2*). In *lane 3* cell extracts were prepared in the presence of 5 mM dithiothreitol (*DTT*). Notice the decrease of the 100-250-kDa aggregates and the slightly increased intensity of the 45-kDa band in *lane 2* as compared to *lanes 1* and *3*. Samples were resolved on a 4-20% SDS-PAGE.

[View Larger Version of this Image (28K GIF file)]

Phosphorylation of the Presenilins

The high level of serines, threonines, and tyrosines in the primary amino acid sequences of both PS1 and PS2 suggested the possibility that the PS proteins are phosphoproteins (12, 13). Transfected COS-1 cells

(Fig. 7) and CHO cells (Fig. 8*B*) incorporated ^{32}P in presenilins. Acid hydrolysis of immunoprecipitated ^{32}P-phosphorylated PS1 and PS2 yielded mainly phosphoserine (Fig. 7*B*). Longer exposures revealed very weak signals for phosphothreonine, while phosphotyrosine was never observed. The phosphorylation of PS1 was strongly increased, and that of PS2 slightly or not increased, by treating the cells with the phosphatase inhibitor okadaic acid (300 nM, Fig. 8*A*). Okadaic acid inhibits the two major classes of serine/threonine phosphatases (49). The protein kinase C inhibitor staurosporin (300 nM) or the agonists phorbol myristic acid (1 μM) and phorbol dibutyric acid (1 μM) and the protein kinase A agonist forskolin (100 μM) had no effect on the phosphorylation extent or pattern (Fig. 8*A*). Attempts to phosphorylate immunoprecipitated PS1 *in vitro* using protein kinase A or protein kinase C remained also negative (results not shown).

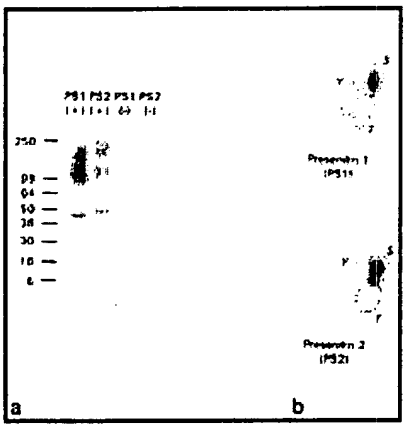


Fig. 7. Phosphorylation of PS1 and PS2 on serine residues. *Panel a*, transfected COS-1 cells were metabolically labeled using [^{32}P]orthophosphate and post-nuclear cell extracts were prepared (see "Materials and Methods"). PS1 and PS2 were immunoprecipitated with antiserum B16 (PS1) or 519 (PS2). COS-1 cells transfected with pSG5 were labeled, extracted, and immunoprecipitated with the same antibodies (−). *Panel b*, thin layer chromatography of hydrolyzed PS1 (45 kDa) and PS2 (50 kDa) identifying serine as the phosphorylated residues. The mobility of phosphoserine (*S*), phosphothreonine (*T*), or phosphotyrosine (*Y*), as detected by ninhydrin staining, is indicated by the *dotted lines*.
[View Larger Version of this Image (31K GIF file)]

Fig. 8. Analysis of PS1 and PS2 phosphorylation. *Panel A*, staurosporin (300 nM) or okadaic acid (300 nM) were added to the medium of transfected COS cells after a 3.5-h incubation with [^{32}P]orthophosphate. 30 min later, cell extracts were made as in Fig. 4. PS1 or PS2 were immunoprecipitated using, respectively, antiserum B17 or 519 and resolved by a 4-20% gradient gel. *Panel B*, transfected CHO cells were labeled with [^{32}P]orthophosphate and PS was immunoprecipitated using antiserum 519 (*PS2 WT*) or B16 (all other lanes) and resolved by a 4-20% gradient gel.
[View Larger Version of this Image (30K GIF file)]

Cytoplasmic Orientation of the Amino-terminal Domain and Hydrophilic Loop Domain of Presenilin 1

We finally addressed the issue whether the two major hydrophilic domains in presenilin 1, *i.e.* the amino-terminal domain and the hydrophilic loop domain (12) are oriented toward the cytoplasmic or the luminal side of the endoplasmic reticulum. Current models for the orientation of the presenilins in membranes are based on theoretical predictions (see "Discussion"). Both domains are candidate regions for

interactions with other proteins, and their orientation determines whether this interaction occurs with cytoplasmic or endoplasmic reticulum proteins. To analyze this question, we used low concentrations of digitonin to selectively permeabilize the plasma membrane (see "Materials and Methods") and antibodies against the amino-terminal domain (antiserum B14), against the Myc tag introduced at the amino terminus (mAb 9E10), against the second loop domain (antibody 519), and, finally, against the hydrophilic loop domain (antiserum B16) of PS1. To monitor the permeabilization procedure, rat monoclonal antibody against the luminal endoplasmic reticulum retention signal KDEL was used (30, 31). Antibodies against the amino-terminal domain of PS1, against the amino-terminally inserted Myc tag (Fig. 9*a*) or against the hydrophilic loop (Fig. 9*e*) reacted with PS1 in digitonin-permeabilized cells under conditions in which the KDEL-specific monoclonal antibody did not result in labeling (Fig. 9, *a* and *e*). On the other hand, antibodies against the second hydrophilic loop domain (519) reacted with PS1 when saponin (Fig. 9*d*), but not when digitonin (Fig. 9*c*) was used to permeabilize the cells, suggesting a luminal localization of this domain.

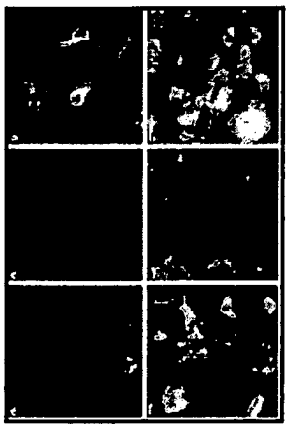


Fig. 9. Cytoplasmic orientation of the amino-terminal domain and the hydrophilic loop of PS1. COS-1 cells transfected with PS1Myc (*panels a* and *b*) or with PS1 WT (*panels c-f*) were permeabilized using digitonin (*panels a*, *c*, and *e*) or saponin (*panels b*, *d*, and *f*) and immunostained with anti-Myc mAb 9E10 (*panels a* and *b*), with polyclonal antibody 519 against the second loop domain (*panels c* and *d*), with polyclonal antibody B16 (*panels e* and *f*) and with rat monoclonal antibody against the KDEL sequence (*panels a-f*). PS1 immunoreactivity is observed as *red/yellow* staining, while KDEL immunoreactivity is *green*. *Panels a* and *b* show that the amino terminus of PS1 is oriented to the cytoplasmic side of the endoplasmic reticulum membrane. Cells that express PS1Myc stain strongly with the amino-terminal domain directed 9E10 mAb both in digitonin (*a*) or saponin (*b*) permeabilized cells. When saponin is used, the untransfected cells are stained with the KDEL antibody (*green*) since it reacts with endogenous proteins (*b*). When digitonin is used, no staining is observed (*a*), indicating that the endoplasmic reticulum membrane is not permeable for antibodies under the conditions used. In *panels c* and *d*, similar experiments are displayed but using antiserum 519 against the second loop domain. Staining is only observed when saponin is used, indicating a luminal orientation of this domain (*panel d*). *Panels e* and *f* demonstrate that antiserum B16 against the hydrophilic loop domain of PS1 react both with digitonin (*e*) or saponin (*f*) permeabilized cells, indicating a cytoplasmic orientation of the hydrophilic loop.
[View Larger Version of this Image (91K GIF file)]

DISCUSSION

The current investigation provides a detailed characterization of the biosynthesis of presenilins overexpressed in COS-1 cells and CHO cells. The results show that the PS proteins are unglycosylated phosphoproteins with apparent molecular masses of about 45 and 50 kDa in SDS-PAGE. They have a

strong tendency to form SDS-resistant complexes with apparent molecular masses between 100 and 250 kDa. Double immunofluorescence studies showed that PS1 is mainly located in the endoplasmic reticulum and the Golgi apparatus and that its amino-terminal domain and the hydrophilic loop domain are oriented to the cytoplasmic side.

The localization of transfected PS1 in the endoplasmic reticulum of COS-1 cells or CHO cells is not a simple consequence of overexpression of the protein or nonspecific general effects on the biosynthetic pathway in these cells. First, endogeneous presenilins are located in the endoplasmic reticulum of untransfected cells as demonstrated in Fig. 1*A* using confocal laser microscopy. Second, using identical transfection conditions, APP and furin were found to accumulate in the Golgi apparatus, as demonstrated before (33, 34). Third, the turnover (Fig. 4) and the secretory processing of APP[3] in CHO cells stably expressing human APP770 was not affected by transfection of PS1, ruling out general inhibitory effects on protein transport and processing. On the other hand, the levels of endogeneous presenilin in COS-1 cells are apparently very low. Only the combination of immunoprecipitation to concentrate PS1 from detergent extracts of 10×10^6 cells, followed by immunoblotting using mAb PS1-3, allowed us to detect endogeneous PS1 migrating at the same molecular weight as transfected PS1. The disadvantage of this approach is that the antibodies from the immunoprecipitation step interfere with the consecutive detection of possible PS-aggregates in the immunoblotting step (Fig. 3*C*). This assay also precludes a dynamic analysis of the metabolism of the presenilins using metabolic labeling and pulse-chase experiments. On the other hand, it allowed us to demonstrate the presence or absence of particular antibody epitopes on PS1 fragments. Importantly, the use of hydrophilic loop-specific antibodies in the immunoprecipitation step resulted in the detection of relative large amounts of, presumably proteolytic, fragments of PS1 in the 18-30-kDa range (Fig. 3*C*). The level of these fragments was not, or only slightly, increased in COS-1 cells that overexpressed PS1, which explains why they were not readily detected in the transfection experiments. The proteolytic process involved is probably easily saturated or even inhibited by PS overexpression (35, 36). Our data therefore do not allow us to speculate any further on the exact nature or the biological significance of this process. We can only conclude that overexpression studies of the type used here are not suitable to study this particular aspect of presenilin metabolism. On the other hand, it is clear that untransfected cells (Fig. 3) and brain tissue *in vivo* (36, 37) contain detectable amounts of intact presenilin. Since this endogeneously expressed PS1 has the same mobility as transfected PS1, the conclusion that transfected PS1 (and PS2) are not subject to glycosylation, sulfation, glycosaminoglycan modification, palmitoylation, or myristoylation also holds apparently true for the endogeneous protein. These posttranslational modifications are thereby also excluded as contributing to the formation of the high molecular mass, SDS-resistant aggregates (100-250 kDa) in SDS-PAGE electrophoresis. These aggregates were observed to a variable extent in all our experiments, and were noticed by others using other cell types or brain tissue (36, 38, 39). Since the smears were detected with three different antibodies recognizing three different epitopes, and by a Myc-directed mAb using a Myc-tagged PS1 construct, the aggregates must contain presenilin core proteins, alone or associated with unidentified components. Denaturation of immunoprecipitates at 37 °C instead of 95 °C, resulted in less aggregates and increased amounts of the 45-kDa PS1 band (Fig. 3*C*). The aggregates therefore probably consist mainly of oligomers of presenilins. It is unclear whether this property to form SDS-resistant aggregates *in vitro* has any physiological significance *in vivo*, but immuno-electron microscope observations suggest that clustering of presenilins also occurs in the endoplasmic reticulum of untransfected cells.[2] It should be envisaged that in pathological

conditions exacerbated aggregation of presenilins could become a problem. In this context, it should be mentioned that antibodies against PS1 stain amyloid plaques in the brains of Alzheimer's disease patients (40).

Importantly, the two point mutations causing Alzheimer's disease that were studied here, *i.e.* PS1 Ala-246 → Glu (FAD1 kindred) and PS1 Cys-410 → Tyr (NIH2 kindred), did not alter the biosynthesis or level of expression of the PS1 protein in COS-1 cells and CHO cells. These mutations must cause therefore more subtle effects. The possibility that they interfere with the fragmentation of PS1, as was shown recently for two other mutations (35, 36), cannot be excluded from the current data. Analysis of brains of transgenic mice expressing the PS1 Ala-246 → Glu mutant indicate, however, that at least this mutation does not interfere with the proteolytic processing of PS1.[4] Since all PS mutations known to date act in a dominant way in the pathogenesis of Alzheimer's disease, a gain of function is most likely. Evidence that fibroblasts derived from patients with presenilin mutations produce more and longer forms of the $^{\beta}$A4 amyloid peptide suggests that the mutations exert their effect on cellular metabolism or trafficking of amyloid precursor protein (21). The localization of PS1 in the early compartments of the biosynthetic pathway makes it unlikely that the presenilin mutations directly influence α- and/or $^{\beta}$-secretase activity, since these operate in the late-Golgi and transport vesicles, at the cell surface and in endosomes (27, 41-43). The possibility that PS mutants influence the balance between amyloidogenic and non-amyloidogenic processing of APP in an indirect way by changing its intracellular trafficking should now be further explored in polarized Madin-Darby canine kidney cells and neurons (16, 44, 45).

Our study demonstrates furthermore that the presenilins are phosphorylated on serine residues. Phosphorylation of PS1 and PS2 was evident both in COS-1 cells and CHO cells, but the phosphorylation of PS1 was less intense and more variable than of PS2. The observation that okadaic acid enhanced the PS1 phosphorylation suggests that PS1 is more prone to phosphatase activity than PS2 (Fig. 8*A*). Protein kinase C is not responsible for PS phosphorylation, since neither stimulation of cells by phorbol esters nor purified kinase added to immunoprecipitated PS1 or PS2 resulted in increased labeling. It is therefore unlikely that the presenilins are directly involved in the regulation of APP secretion by protein kinase C (46, 47).

Finally, we showed that the hydrophilic NH_2-terminal and the major loop domain of PS1 are exposed to the cytoplasmic side of the endoplasmic reticulum membrane. Digitonin at low concentrations selectively permeabilizes the cell membrane, but not the endoplasmic reticulum membrane (29). Antibodies directed toward the amino-terminal domain of PS1 (Fig. 9) displayed similar immunofluorescent staining patterns in digitonin and saponin permeabilized cells. Control experiments with antibodies against a luminal epitope, *i.e.* the KDEL endoplasmic reticulum retention signal (30, 31), showed that digitonin did not permeabilize the endoplasmic reticulum membrane. For the major hydrophilic loop domain, essentially identical results were obtained. The conclusion therefore is that the two major hydrophilic domains in PS1 are oriented to the cytoplasmic side of the endoplasmic reticulum, which directs the search for candidate proteins interacting with these domains toward the cytoplasm. Consistent with this conclusion, antibodies against the second loop domain reacted with PS1 after saponin, but not after digitonin permeabilization. This not only confirms the luminal localization of the recognized epitope, but also validates the approach as it was used.

While more detailed studies, including electron microscopy and enzymatic digestion protection assays, are

required to settle definitively the orientation of the presenilins in the endoplasmic reticulum membrane, our results clarify already two important issues. The first problem is the orientation of the first transmembrane domain in PS1, which determines in fact the topology of all the following transmembrane domains (46). The positive charge difference between the 15 carboxyl-terminal and the 15 amino-terminal residues flanking the first membrane spanning domain suggests a luminal orientation of its amino terminus (48). However, several exceptions to this rule are known, mainly of proteins of which the amino-terminal domain contains more then 17 charged amino acid residues (48). In PS1, this domain contains 29 charged residues, which would explain its cytoplasmic orientation. The second issue, which has been pointed out by others before (9, 13, 17), is whether two putative hydrophobic stretches in the PS amino acid sequence flanking the hydrophilic loop, can additionally span the membrane, which would make the presenilins nine transmembrane domain proteins. Our data are compatible with a seven-transmembrane domain model with the amino-terminal domain and the hydrophilic loop located in the cytoplasm (Fig. 10; see also *Note Added in Proof*). In conclusion, the current study has analyzed the biosynthesis and the subcellular localization of the presenilins and provides a basis for the further study of their cell biology and their possible interactions with integral membrane and cytoplasmic proteins such as APP or Tau, both implicated in the pathogenesis of Alzheimer's disease (4).

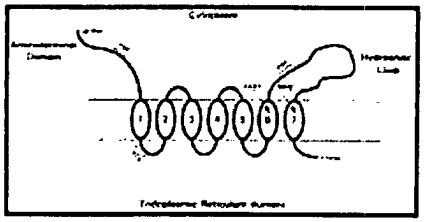


Fig. 10. Localization of antibody epitopes and orientation of PS1 in the endoplasmic reticulum membrane. PS1 is inserted in the endoplasmic reticulum membrane with both its amino-terminal (NH_2-) and hydrophilic loop domain at the cytoplasmic side. The localization of the peptides used to raise the different antibodies used to study the topology of PS1 are indicated by *bars* above the sequence (see also "Materials and Methods" and *Note Added in Proof*).
[View Larger Version of this Image (13K GIF file)]

FOOTNOTES

* This investigation was supported by grants from the Fonds voor Geneeskundig Wetenschappelijk Onderzoek (FGWO-NFWO), the Human Frontiers of Science Program (HFSP), the Action Program for Biotechnology of the Flemish Government (VLAB), the Flemish Institute for Biotechnology (VIB), the Katholieke Universiteit Leuven, the Alzheimer's Society of Ontario, and the Medical Research Council of Canada. The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Onderzoeksleider of the National Fund for Scientific Research, Belgium (NFWO). To whom correspondence should be addressed: Campus Gasthuisberg, O&N 06, 3000 Leuven, Belgium. Fax: 32-16-345871; E-mail: bart.destrooper@med.kuleuven.ac.be.

[1] The abbreviations used are: PS2, presenilin 2; PS1, presenilin 1; APP, amyloid precursor protein; mAb, monoclonal antibody; PBS, phosphate-buffered saline; TBS, Tris-buffered saline; FITC, fluorescein

isothiocyanate; TRITC, tetramethylrhodamine isothiocyanate; PAGE, polyacrylamide gel electrophoresis; Pipes, 1,4-piperazinediethanesulfonic acid; NSP, neuroendocrine-specific protein.

[2] P. Fraser, unpublished results.

[3] B. De Strooper and K. Craessaerts, unpublished results.

[4] D. Moechars, B. De Strooper, and F. Van Leuven, unpublished results.

Acknowledgments

We gratefully acknowledge the contribution of expertise or materials of the following scientists: B. Greenberg; A. Roebroek, J. Creemers, and W. Van De Ven; and G. Schellenberg (Geriatric Research, Education and Clinical Center, Seattle, WA).

Note Added in Proof

An eight-transmembrane model, with hydrophobic region VIII in the loop domain spanning the endoplasmic reticulum membrane, is also compatible with these data. This would result in a cytoplasmic orientation of the carboxyl-terminal domain of PS1.

REFERENCES

1. Haass, C., and Selkoe, D. J. (1993) *Cell* **75**, 1039-1042 [Medline]
2. Selkoe, D. J. (1994) *Annu. Rev. Cell Biol.* **10**, 373-403
3. Checler, F. (1995) *J. Neurochem.* **65**, 1431-1444 [Medline]
4. Kosik, K. S. (1994) *J. Cell Biol.* **127**, 1501-1504 [Medline]
5. Haass, C. (1996) *Curr. Opin. Neurol.* **9**, 254-259 [Medline]
6. Strittmatter, W. J., and Roses, A. D. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 4725-4727 [Abstract]
7. Rebeck, G. W., Reiter, J. S., Strickland, D. K., and Hyman, B. T. (1993) *Neuron* **11**, 575-580 [Medline]
8. Goate, A., Chartier-Harlin, M. C., Mullan, M., Brown, J., Crawford, F., Fidani, L., Giuffra, L., Haynes, A., Irving, N., James, L., Mant, R., Newton, P., Rooke, K., Roques, P., Talbot, C., Pericak-Vance, M., Roses, A., Williamson, Rossor, M., Owen, M., and Hardy, J. (1991) *Nature* **349**, 704-706 [Medline]
9. Rogaev, E. I., Sherrington, R., Rogaeva, E. A., Levesque, G., Ikeda, M., Liang, Y., Chi, H., Lin, C., Holamn, K., Tsuda, T., Mar, L., Sorbi, S., Nacmias, B., Piacenti, S., Amaducci, L., Chumakov, I., Cohen, D., Lannfelt, L., Fraser, P. E., Rommens, J. M., and St. George-Hyslop, P. H. (1995) *Nature* **376**, 775-778 [Medline]
10. Levy-Lahad, E., Wasco, W., Poorkaj, P., Romano, D. M., Osima, J., Pettingell, H., Yu, C., Jondro, P. D., Schmidt, S. D., Wang, K., Crowley, A. C., Fu, Y. H., Guenette, S. Y., Galas, D., Nemens, E., Wijsman, E. M., Bird, T. D., Schellenberg, G. D., and Tanzi, R. E. (1995) *Science* **269**, 973-977 [Medline]
11. Li, J., Ma, J., and Potter, H. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 12180-12184 [Abstract]
12. Sherrington, R., Rogaev, E. I., Liang, Y., Rogaeva, E. A., Levesque, G., Ikeda, M., Chi, H., Lin, C., Li, G., Holman, K., Tsuda, T., Mar, L., Foncin, J. F., Bruni, A. C., Montesi, M. P., Sorbi, S.,

Rainero, I., Pinessi, L., Nee, L., Chumakov, I., Pollen, D., Brookes, A., Sanseau, P., Polnsky, R. J., Wasco, W., da Silva, H. A. R., Haines, J. L., Pericak-Vance, M. A., Tanzi, R. E., Roses, A. D., Fraser, P. E., Rommens, J. M., and St. George-Hyslop, P. H. (1995) *Nature* **375**, 754-760 [Medline]

13. Alzheimer's Disease Collaborative Group (1995) *Nat. Genet.* **11**, 219-222 [Medline]
14. Wasco, W., Pettingell, W. P., Jondro, P. D., Schmidt, S. D., Gurubhagavatula, S., Rodes, L., DiBlasi, T., Romano, T. M., Guenette, S. Y., Kovacs, D. M., Growdon, J. H., and Tanzi, R. E. (1995) *Nat. Med.* **1**, 848
15. Van Broeckhoven, C. (1995) *Nat. Genet.* **11**, 230-232 [Medline]
16. De Strooper, B., Simons, M., Multhaup, G., Van Leuven, F., Beyreuther, K., and Dotti, C. G. (1995) *EMBO J.* **14**, 4932-4938 [Abstract]
17. Slunt, H. H., Thinakaran, G., Lee, M. K., and Sisodia, S. S. (1995) *Int. J. Exp. Clin. Invest.* **2**, 188-190
18. Pasquale, V., Lacana, E., and D'Adamio, L. (1996) *Science* **271**, 521-525 [Abstract]
19. L'Hernault, S. W., and Arduengo, P. M. (1992) *J. Cell Biol.* **119**, 55-68 [Abstract]
20. Leviatan, D., and Greenwald, I. (1995) *Nature* **377**, 351-354 [Medline]
21. Scheuner, D., Eckman, C., Jensen, M., Song, X., Citron, M., Suzuki, N., Bird, T. D., Hardy, J., Hutton, M., Kukull, W., Larson, E., Levy-Lahad, E., Viitanen, M., Peskind, E., Poorkaj, P., Schellenberg, G., Tanzi, R., Wasco, W., Lannfelt, L., Selkoe, D., and Younkin, S. (1996) *Nat. Med.* **2**, 864-870 [Medline]
22. Creemers, J. W. M., Vey, M., Schäfer, W., Ayoubi, T. A. Y., Roebroek, A. J. M., Klenk, H.-D., Garten, W., and Van de Ven, W. J. M. (1995) *J. Biol. Chem.* **270**, 2695-2702 [Abstract/Full Text]
23. Roebroek, A. J. M., Schalken, A., Leunissen, J. A. M., Onnekink, C., Bloemers, H. P. J., and Van de Ven, W. J. M. (1986) *EMBO J.* **5**, 2197-2202 [Abstract]
24. Van de Velde, H. J. K., Roebroek, A. J. M., Senden, N. H. M., Ramaekers, F. C. S., and Van De Ven, W. J. M. (1994) *J. Cell Sci.* **107**, 2403-2416 [Abstract]
25. Roebroek, A. J. M., Van de Velde, H. J. K., Van Bokhoven, A., Broers, J. L. V., Ramaekers, F. C. S., and Van de Ven, W. J. M. (1993) *J. Biol. Chem.* **268**, 13439-13477 [Abstract]
26. Evan, G., Lewis, G. K., Ramsay, G., and Bishop, J. M. (1985) *Mol. Cell. Biol.* **5**, 3610-3616 [Medline]
27. De Strooper, B., Umans, L., Van Leuven, F., and Van Den Berghe, H. (1993) *J. Cell Biol.* **121**, 295-304 [Abstract]
28. Cooper, J. A., Sefton, B. M., and Hunter, T. (1983) *Methods Enzymol.* **99**, 387-402 [Medline]
29. Otto, J. C., and Smith, W. L. (1994) *J. Biol. Chem.* **269**, 19868-19875 [Abstract]
30. Napier, R. M., Venis, M. A., Bolton, M. A., Richardson, L. I., and Butcher, G. W. (1988) *Planta* **176**, 519-526
31. Munro, S., and Pelham, H. R. B. (1987) *Cell* **48**, 899-907 [Medline]
32. Jorgensen, A., Arnold, W., Pepper, D., Kahl, S., Mandel, F., and Campbell, K. (1988) *Cell Motil. Cytoskel.* **9**, 164-174 [Medline]
33. Slunt, H. H., Thakaran, G., Von Koch, C., Lo, A. C. Y., Tanzi, R. E., and Sisodia, S. S. (1994) *J. Biol. Chem.* **269**, 2637-2644 [Abstract]
34. De Strooper, B., Craessaerts, K., Van Leuven, F., and Van Den Berghe, H. (1995) *J. Biol. Chem.* **270**, 30310-30314 [Abstract/Full Text]
35. Thinakaran, G., Borchelt, D. R., Lee, M. K., Slunt, H. H., Spitzer, L., Kim, G., Ratovitsky, T., Davenport, F., Nordstedt, C., Seeger, M., Hardy, J., Levey, A. I., Gandy, S. E., Jenkins, N. A., Copeland, N. G., Price, D. L., and Sisodia, S. S. (1996) *Neuron* **17**, 181-190 [Medline]
36. Mercken, M., Takahashi, H., Honda, T., Sato, K., Muryama, M., Nakazato, Y., Noguchi, K., Imahori, K., and Takashima, A. (1996) *FEBS Lett.* **389**, 297-303 [CrossRef][Medline]
37. Elder, G. A., Tezapsidis, N., Carter, J., Shioi, J., Bouras, C., Li, H. C., Johnston, J. M., Efthimiopoulos, S., Friedrich, V. L., and Robakis, N. K. (1996) *J. Neurosci. Res.* **45**, 308-320 [CrossRef][Medline]

38. Kovacs, D. M., Fausett, H., Page, K. J., Kim, T. W., Moir, R. D., Merriam, D. E., Hollister, R. D., Hallmark, O. G., Mancini, R., Felsenstein, K. M., Hyman, B. T., Tanzi, R. E., and Wasco, W. (1996) *Nat. Med.* **2**, 1-6
39. Moussaoui, S., Czech, C., Pradier, L., Blanchard, V., Bonici, B., Gohin, M., Imperato, A., and Revah, F. (1996) *FEBS Lett.* **383**, 219-222 [CrossRef][Medline]
40. Wisniewski, T., Palha, J. A., Ghiso, J., and Frangione, B. (1995) *Lancet* **346**, 1366
41. Koo, E. H., and Squazzo, S. L. (1994) *J. Biol. Chem.* **269**, 17386-17389 [Abstract]
42. Thinakaran, G., Teplow, D. B., Siman, R., Greenberg, B., and Sisodia, S. S. (1996) *J. Biol. Chem.* **271**, 9390-9397 [Abstract/Full Text]
43. Haass, C., Lemere, C. A., Capell, A., Citron, M., Seubert, P., Schenk, D., Lannfelt, L., and Selkoe, D. (1995) *Nat. Med.* **1**, 1291-1296 [Medline]
44. Haass, C., Koo, E. H., Capell, A., Teplow, D. B., and Selkoe, D. J. (1995) *J. Cell Biol.* **128**, 537-547 [Abstract]
45. De Strooper, B., Craessaerts, K., Dewachter, I., Moechars, D., Greenberg, B., Van Leuven, F., and Van Den Berghe, H. (1995) *J. Biol. Chem.* **270**, 4058-4065 [Abstract/Full Text]
46. Buxbaum, J. D., Gandy, S. E., Cichetti, P., Ehrlich, M. E., Czernik, A. J., Fracasso, R. P., Ramabhadran, T. V., Unterbeck, A. J., and Greenhard, P. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 6003-6006 [Abstract]
47. Gillespie, S. L., Golde, T. E., and Younkin, S. G. (1992) *Biochem. Biophys. Res. Commun.* **187**, 1285-1290 [Medline]
48. Hartmann, E., Rapoport, T. A., and Lodish, H. F. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 5786-579049 [Medline]
49. Bollen, M., and Stalmans, W. (1992) *Crit. Rev. Biochem. Mol. Biol.* **27**, 227-281 [Medline]
50. Lories, V., Cassiman, J.-J., Van den Berghe, H., and David, G. (1989) *J. Biol. Chem.* **264**, 7009-7016 [Abstract]

This article has been cited by other articles:

- Dewji, N. N., Singer, S. J. (1997). Cell surface expression of the Alzheimer disease-related presenilin proteins. *Proc. Natl. Acad. Sci. U. S. A.* 94: 9926-9931 [Abstract] [Full Text]
- Keller, J. N., Guo, Q., Holtsberg, F. W., Bruce-Keller, A. J., Mattson, M. P. (1998). Increased Sensitivity to Mitochondrial Toxin-Induced Apoptosis in Neural Cells Expressing Mutant Presenilin-1 Is Linked to Perturbed Calcium Homeostasis and Enhanced Oxyradical Production. *J. Neurosci.* 18: 4439-4450 [Abstract] [Full Text]
- Smine, A., Xu, X., Nishiyama, K., Katada, T., Gambetti, P., Yadav, S. P., Wu, X., Shi, Y.-C., Yasuhara, S., Homburger, V., Okamoto, T. (1998). Regulation of Brain G-protein Go by Alzheimer's Disease Gene Presenilin-1. *J. Biol. Chem.* 273: 16281-16288 [Abstract] [Full Text]
- Steiner, H., Capell, A., Pesold, B., Citron, M., Kloetzel, P. M., Selkoe, D. J., Romig, H., Mendla, K., Haass, C. (1998). Expression of Alzheimer's Disease-associated Presenilin-1 Is Controlled by Proteolytic Degradation and Complex Formation. *J. Biol. Chem.* 273: 32322-32331 [Abstract] [Full Text]

- Abstract of this Article
- **Reprint (PDF) Version of this Article**
- Similar articles found in:
  JBC Online
  PubMed
- PubMed Citation
- This Article has been cited by:
- Search Medline for articles by:
  De Strooper, B. || Van Leuven, F.
- Alert me when:
  new articles cite this article
- Download to Citation Manager

- Walter, J., Schindzielorz, A., Grunberg, J., Haass, C. (1999). Phosphorylation of presenilin-2 regulates its cleavage by caspases and retards progression of apoptosis. *Proc. Natl. Acad. Sci. U. S. A.* 96: 1391-1396 [Abstract] [Full Text]
- Pastorcic, M., Das, H. K. (1999). An Upstream Element Containing an ETS Binding Site Is Crucial for Transcription of the Human Presenilin-1 Gene. *J. Biol. Chem.* 274: 24297-24307 [Abstract] [Full Text]
- Price, D. L., Sisodia, S. S. (1998). MUTANT GENES IN FAMILIAL ALZHEIMER'S DISEASE AND TRANSGENIC MODELS. *Annu. Rev. Neurosci.* 21: 479-505 [Abstract] [Full Text]
- Guo, Y., Livne-Bar, I., Zhou, L., Boulianne, G. L. (1999). Drosophila presenilin Is Required for Neuronal Differentiation and Affects Notch Subcellular Localization and Signaling. *J. Neurosci.* 19: 8435-8442 [Abstract] [Full Text]
- Annaert, W. G., Levesque, L., Craessaerts, K., Dierinck, I., Snellings, G., Westaway, D., George-Hyslop, P. St., Cordell, B., Fraser, P., De Strooper, B. (1999). Presenilin 1 Controls {gamma}-Secretase Processing of Amyloid Precursor Protein in Pre-Golgi Compartments of Hippocampal Neurons. *J. Cell Biol.* 147: 277-294 [Abstract] [Full Text]
- Donoviel, D. B., Hadjantonakis, A.-K., Ikeda, M., Zheng, H., Hyslop, P. S. G., Bernstein, A. (1999). Mice lacking both presenilin genes exhibitearlyembryonic patterningdefects. *Genes & Dev.* 13: 2801-2810 [Abstract] [Full Text]
- Sudoh, S., Hua, G., Kawamura, Y., Maruyama, K., Komano, H., Yanagisawa, K. (2000). Intracellular site of {gamma}-secretase cleavage for A{beta}42 generation in Neuro 2a cells harbouring a presenilin 1 mutation. *Eur J Biochem* 267: 2036-2045 [Abstract] [Full Text]
- Walter, J., Grunberg, J., Capell, A., Pesold, B., Schindzielorz, A., Citron, M., Mendla, K., George-Hyslop, P. S., Multhaup, G., Selkoe, D. J., Haass, C. (1997). Proteolytic processing of the Alzheimer disease-associated presenilin-1 generates an in vivo substrate for protein kinase C. *Proc. Natl. Acad. Sci. U. S. A.* 94: 5349-5354 [Abstract] [Full Text]
- Herreman, A., Hartmann, D., Annaert, W., Saftig, P., Craessaerts, K., Serneels, L., Umans, L., Schrijvers, V., Checler, F., Vanderstichele, H., Baekelandt, V., Dressel, R., Cupers, P., Huylebroeck, D., Zwijsen, A., Van Leuven, F., De Strooper, B. (1999). Presenilin 2 deficiency causes a mild pulmonary phenotype and no changes in amyloid precursor protein processing but enhances the embryonic lethal phenotype of presenilin 1 deficiency. *Proc. Natl. Acad. Sci. U. S. A.* 96: 11872-11877 [Abstract] [Full Text]
- Shirotani, K., Takahashi, K., Araki, W., Maruyama, K., Tabira, T. (2000). Mutational Analysis of Intrinsic Regions of Presenilin 2 That Determine Its Endoproteolytic Cleavage and Pathological Function. *J. Biol. Chem.* 275: 3681-3686 [Abstract] [Full Text]
- Parkin, E. T., Hussain, I., Karran, E. H., Turner, A. J., Hooper, N. M. (1999). Characterization of Detergent-Insoluble Complexes Containing the Familial Alzheimer's Disease-Associated Presenilins. *J Neurochem* 72: 1534-1543 [Abstract] [Full Text]
- Begley, J. G., Duan, W., Chan, S., Duff, K., Mattson, M. P. (1999). Altered Calcium Homeostasis and Mitochondrial Dysfunction in Cortical Synaptic Compartments of Presenilin-1 Mutant Mice. *J Neurochem* 72: 1030-1039 [Abstract] [Full Text]
- Levesque, G., Yu, G., Nishimura, M., Zhang, D. M., Levesque, L., Yu, H., Xu, D., Liang, Y., Rogaeva, E., Ikeda, M., Duthie, M., Murgolo, N., Wang, L., VanderVere, P., Bayne, M. L., Strader, C. D., Rommens, J. M., Fraser, P. E., George-Hyslop, P. St. (1999). Presenilins Interact with Armadillo Proteins Including Neural-Specific Plakophilin-Related Protein and {beta}-Catenin. *J Neurochem* 72: 999-1008 [Abstract] [Full Text]
- Pinnix, I., Musunuru, U., Tun, H., Sridharan, A., Golde, T., Eckman, C., Ziani-Cherif, C., Onstead, L., Sambamurti, K. (2000). A Novel gamma -Secretase Assay Based on Detection of the Putative C-terminal Fragment-gamma of Amyloid beta Protein Precursor. *J. Biol. Chem.* 276: 481-487 [Abstract] [Full Text]
- Counts, S. E., Lah, J. J., Levey, A. I. (2001). The regulation of presenilin-1 by nerve growth factor. *J*

Neurochem 76: 679-689 [Abstract] [Full Text]

- Soriano, S., Kang, D. E., Fu, M., Pestell, R., Chevallier, N., Zheng, H., Koo, E. H. (2001). Presenilin 1 Negatively Regulates {beta}-Catenin/T Cell Factor/Lymphoid Enhancer Factor-1 Signaling Independently of {beta}-Amyloid Precursor Protein and Notch Processing. *J. Cell Biol.* 152: 785-794 [Abstract] [Full Text]
- Schneider, I., Reverse, D., Dewachter, I., Ris, L., Caluwaerts, N., Kuiperi, C., Gilis, M., Geerts, H., Kretzschmar, H., Godaux, E., Moechars, D., Van Leuven, F., Herms, J. (2001). Mutant Presenilins Disturb Neuronal Calcium Homeostasis in the Brain of Transgenic Mice, Decreasing the Threshold for Excitotoxicity and Facilitating Long-term Potentiation. *J. Biol. Chem.* 276: 11539-11544 [Abstract] [Full Text]
- Kim, J. W., Chang, T.-S., Lee, J. E., Huh, S.-H., Yeon, S. W., Yang, W. S., Joe, C. O., Mook-Jung, I., Tanzi, R. E., Kim, T.-W., Choi, E.-J. (2001). Negative Regulation of the SAPK/JNK Signaling Pathway by Presenilin 1. *J. Cell Biol.* 153: 457-464 [Abstract] [Full Text]
- Wild-Bode, C., Yamazaki, T., Capell, A., Leimer, U., Steiner, H., Ihara, Y., Haass, C. (1997). Intracellular Generation and Accumulation of Amyloid beta -Peptide Terminating at Amino Acid 42. *J. Biol. Chem.* 272: 16085-16088 [Abstract] [Full Text]
- Johnston, J. A., Ward, C. L., Kopito, R. R. (1998). Aggresomes: A Cellular Response to Misfolded Proteins. *J. Cell Biol.* 143: 1883-1898 [Abstract] [Full Text]
- Hong, C.-S., Caromile, L., Nomata, Y., Mori, H., Bredesen, D. E., Koo, E. H. (1999). Contrasting Role of Presenilin-1 and Presenilin-2 in Neuronal Differentiation In Vitro. *J. Neurosci.* 19: 637-643 [Abstract] [Full Text]
- Kang, D. E., Soriano, S., Frosch, M. P., Collins, T., Naruse, S., Sisodia, S. S., Leibowitz, G., Levine, F., Koo, E. H. (1999). Presenilin 1 Facilitates the Constitutive Turnover of beta -Catenin: Differential Activity of Alzheimer's Disease-Linked PS1 Mutants in the beta -Catenin-Signaling Pathway. *J. Neurosci.* 19: 4229-4237 [Abstract] [Full Text]
- Nakai, T., Yamasaki, A., Sakaguchi, M., Kosaka, K., Mihara, K., Amaya, Y., Miura, S. (1999). Membrane Topology of Alzheimer's Disease-related Presenilin 1. EVIDENCE FOR THE EXISTENCE OF A MOLECULAR SPECIES WITH A SEVEN MEMBRANE-SPANNING AND ONE MEMBRANE-EMBEDDED STRUCTURE. *J. Biol. Chem.* 274: 23647-23658 [Abstract] [Full Text]
- Passer, B. J., Pellegrini, L., Vito, P., Ganjei, J. K., D'Adamio, L. (1999). Interaction of Alzheimer's Presenilin-1 and Presenilin-2 with Bcl-XL. A POTENTIAL ROLE IN MODULATING THE THRESHOLD OF CELL DEATH. *J. Biol. Chem.* 274: 24007-24013 [Abstract] [Full Text]
- Octave, J.-N., Essalmani, R., Tasiaux, B., Menager, J., Czech, C., Mercken, L. (2000). The Role of Presenilin-1 in the gamma -Secretase Cleavage of the Amyloid Precursor Protein of Alzheimer's Disease. *J. Biol. Chem.* 275: 1525-1528 [Abstract] [Full Text]
- Lin, X., Koelsch, G., Wu, S., Downs, D., Dashti, A., Tang, J. (2000). Human aspartic protease memapsin 2 cleaves the beta -secretase site of beta -amyloid precursor protein. *Proc. Natl. Acad. Sci. U. S. A.* 97: 1456-1460 [Abstract] [Full Text]
- Dewachter, I., Van Dorpe, J., Smeijers, L., Gilis, M., Kuipéri, C., Laenen, I., Caluwaerts, N., Moechars, D., Checler, F., Vanderstichele, H., Van Leuven, F. (2000). Aging Increased Amyloid Peptide and Caused Amyloid Plaques in Brain of Old APP/V717I Transgenic Mice by a Different Mechanism than Mutant Presenilin1. *J. Neurosci.* 20: 6452-6458 [Abstract] [Full Text]
- Theuns, J., Van Broeckhoven, C. (2000). Transcriptional regulation of Alzheimer's disease genes: implications for susceptibility. *Hum Mol Genet* 9: 2383-2394 [Abstract] [Full Text]
- Jacobsen, H., Reinhardt, D., Brockhaus, M., Bur, D., Kocyba, C., Kurt, H., Grim, M. G., Baumeister, R., Loetscher, H. (1999). The Influence of Endoproteolytic Processing of Familial Alzheimer's Disease Presenilin 2 on Abeta 42 Amyloid Peptide Formation. *J. Biol. Chem.* 274: 35233-35239 [Abstract] [Full Text]
- Chen, F., Yang, D.-S., Petanceska, S., Yang, A., Tandon, A., Yu, G., Rozmahel, R., Ghiso, J.,

Nishimura, M., Zhang, D. M., Kawarai, T., Levesque, G., Mills, J., Levesque, L., Song, Y.-Q., Rogaeva, E., Westaway, D., Mount, H., Gandy, S., St George-Hyslop, P., Fraser, P. E. (2000). Carboxyl-terminal Fragments of Alzheimer beta -Amyloid Precursor Protein Accumulate in Restricted and Unpredicted Intracellular Compartments in Presenilin 1-deficient Cells. *J. Biol. Chem.* 275: 36794-36802 [Abstract] [Full Text]

- Thinakaran, G., Harris, C. L., Ratovitski, T., Davenport, F., Slunt, H. H., Price, D. L., Borchelt, D. R., Sisodia, S. S. (1997). Evidence That Levels of Presenilins (PS1 and PS2) Are Coordinately Regulated by Competition for Limiting Cellular Factors. *J. Biol. Chem.* 272: 28415-28422 [Abstract] [Full Text]
- Zhang, J., Kang, D. E., Xia, W., Okochi, M., Mori, H., Selkoe, D. J., Koo, E. H. (1998). Subcellular Distribution and Turnover of Presenilins in Transfected Cells. *J. Biol. Chem.* 273: 12436-12442 [Abstract] [Full Text]
- Alberici, A., Moratto, D., Benussi, L., Gasparini, L., Ghidoni, R., Gatta, L. B., Finazzi, D., Frisoni, G. B., Trabucchi, M., Growdon, J. H., Nitsch, R. M., Binetti, G. (1999). Presenilin 1 Protein Directly Interacts with Bcl-2. *J. Biol. Chem.* 274: 30764-30769 [Abstract] [Full Text]
- Yu, G., Chen, F., Levesque, G., Nishimura, M., Zhang, D.-M., Levesque, L., Rogaeva, E., Xu, D., Liang, Y., Duthie, M., George-Hyslop, P. H. S., Fraser, P. E. (1998). The Presenilin 1 Protein Is a Component of a High Molecular Weight Intracellular Complex That Contains beta -Catenin. *J. Biol. Chem.* 273: 16470-16475 [Abstract] [Full Text]
- Deguchi, M., Iizuka, T., Hata, Y., Nishimura, W., Hirao, K., Yao, I., Kawabe, H., Takai, Y. (2000). PAPIN. A NOVEL MULTIPLE PSD-95/Dlg-A/ZO-1 PROTEIN INTERACTING WITH NEURAL PLAKOPHILIN-RELATED Armadillo REPEAT PROTEIN/delta -CATENIN AND p0071. *J. Biol. Chem.* 275: 29875-29880 [Abstract] [Full Text]
- Maltese, W. A., Wilson, S., Tan, Y., Suomensaari, S., Sinha, S., Barbour, R., McConlogue, L. (2001). Retention of the Alzheimer's Amyloid Precursor Fragment C99 in the Endoplasmic Reticulum Prevents Formation of Amyloid beta -Peptide. *J. Biol. Chem.* 276: 20267-20279 [Abstract] [Full Text]
- Li, X., Greenwald, I. (1998). Additional evidence for an eight-transmembrane-domain topology for Caenorhabditis elegans and human presenilins. *Proc. Natl. Acad. Sci. U. S. A.* 95: 7109-7114 [Abstract] [Full Text]
- Wu, G., Hubbard, E. J. A., Kitajewski, J. K., Greenwald, I. (1998). Evidence for functional and physical association between Caenorhabditis elegans SEL-10, a Cdc4p-related protein, and SEL-12 presenilin. *Proc. Natl. Acad. Sci. U. S. A.* 95: 15787-15791 [Abstract] [Full Text]
- Ratovitski, T., Slunt, H. H., Thinakaran, G., Price, D. L., Sisodia, S. S., Borchelt, D. R. (1997). Endoproteolytic Processing and Stabilization of Wild-type and Mutant Presenilin. *J. Biol. Chem.* 272: 24536-24541 [Abstract] [Full Text]
- Xia, W., Zhang, J., Perez, R., Koo, E. H., Selkoe, D. J. (1997). Interaction between amyloid precursor protein and presenilins in mammalian cells: Implications for the pathogenesis of Alzheimer disease. *Proc. Natl. Acad. Sci. U. S. A.* 94: 8208-8213 [Abstract] [Full Text]
- Dewji, N. N., Singer, S. J. (1997). The seven-transmembrane spanning topography of the Alzheimer disease-related presenilin proteins in the plasma membranes of cultured cells. *Proc. Natl. Acad. Sci. U. S. A.* 94: 14025-14030 [Abstract] [Full Text]

Volume 271, Number 46, Issue of November 15, 1996 pp. 29422-29426

The Novel Protein-tyrosine Phosphatase PTP20 Is a Positive Regulator of PC12 Cell Neuronal Differentiation*

(Received for publication, August 26, 1996, and in revised form, September 9, 1996)

Naohito Aoki †, **Yumiko Yamaguchi-Aoki** † **and Axel Ullrich** §

From the Department of Molecular Biology, Max-Planck-Institut für Biochemie, Am Klopferspitz 18A, 82152 Martinsried, Germany

ABSTRACT

A novel cytoplasmic protein-tyrosine phosphatase (PTPase) designated PTP20 was isolated from a PC12 cDNA library and shown to positively regulate the differentiation process in PC12 cells. The PTP20 open reading frame of 453 amino acids contains a single tyrosine phosphatase catalytic domain and displays closest homology to members of the PTP-PEST protein-tyrosine phosphatase family. Transient expression of PTP20 in Rat-1 cells resulted in the expression of a 50-kDa protein which exhibited PTPase activity *in vitro*. Expression of the 2.3-kilobase PTP20 mRNA increased during differentiation of nerve growth factor (NGF)-stimulated PC12 cells. Consistent with this observation, stable overexpression of PTP20 in PC12 cells resulted in accelerated neurite formation following NGF treatment. These findings suggest a positive regulatory role of PTP20 in NGF-dependent neuronal differentiation of PC12 cells.

INTRODUCTION

Protein tyrosine phosphorylation has been shown to play a crucial role in regulating fundamental cellular processes including proliferation, differentiation, and tumorigenesis (1). Phosphorylation of tyrosine residues occurs as a cellular response to a variety of stimuli including growth factors, hormones, or cytokines (2). The initial event triggered by these stimuli include an ordered process of activation, dimerization, and autophosphorylation of receptor tyrosine kinases. This in turn initiates an intracellular signaling cascade which involves the phosphorylation of several other key regulatory proteins. The balance of tyrosine phosphorylation is tightly controlled by the opposing activities of protein-tyrosine kinases and protein-tyrosine phosphatases (PTPases)[1] (3).

Since the recent discovery that the tyrosine kinase Trk (tropomyosin receptor kinase) is the receptor for nerve growth factor (NGF) (4, 5, 6), protein tyrosine phosphorylation has been demonstrated to play an important role in neuronal differentiation. Upon NGF stimulation, PC12 cells can differentiate from an endocrine cell phenotype to a sympathetic neuronal phenotype (7). The role of tyrosine kinases in the differentiation process in PC12 cells has been partially elucidated and begins when NGF binds to the extracellular domain of its cognate receptor. This results in stimulation of tyrosine kinase activity and consequent activation of an ordered cascade of phosphorylation events involving $p21^{ras}$, Raf-1 kinase, and mitogen-activated protein (MAP) kinases (8, 9).

Although the importance of protein-tyrosine kinases, like Trk, in neuronal differentiation is well characterized, the involvement of PTPases in neuronal function and differentiation is still unclear. It has been demonstrated that the protein-tyrosine phosphatase inhibitor orthovanadate can inhibit the NGF-induced differentiation of PC12 cells, suggesting that some PTPases are involved in the neural differentiation process (10). To begin to address the importance of PTPases in neuronal differentiation, we attempted to isolate cDNAs which encode PTPases expressed during the differentiation process of PC12 cells. In this paper, we show the molecular cloning of a novel PTPase, named PTP20, and demonstrate that this PTPase might play a positive regulatory role in NGF-induced differentiation of PC12 cells.

EXPERIMENTAL PROCEDURES

PCR Amplification and cDNA Cloning of Rat PTP20

Degenerate oligonucleotide sense and antisense primers were based on consensus sequences for two highly conserved amino acid stretches within the catalytic domains of PTPases: FW*X*M*X*W and HCSAG(S/I/V)G. Random-primed cDNA (up to 50 ng) from PC12 cell RNA was used as a template for PCR. Both sense and antisense primers were added to a 100-μl reaction mixture containing 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2.5 mM $MgCl_2$, 0.01% bovine serum albumin, all four dNTPs (each at 200 μM), 1 unit of *Taq* polymerase (Boehringer Mannheim), and template cDNA. Thirty-five cycles were curried out on a thermal cycler; each cycle involved incubation at 94 °C for 1 min, at 42 °C for 1 min, and 72 °C for 1 min. The PCR products were separated on a 1.5% agarose gel. Fragments of 350-400 bp were excised, subcloned into pBluescript KS(+), and sequenced. The PTP20 PCR fragment was isolated, radioactively labeled by random priming, and used to screen 1×10^6 plaques from a PC12 cDNA library which had been made

using a pool of poly(A)$^+$ RNA from both undifferentiated and differentiated PC12 cells and a λZAPII synthesis kit (Stratagene). Hybridization was performed in a solution containing 50% (v/v) formamide, 5 × SSC, 5 × Denhardt's solution, 0.05 M sodium phosphate, 1 mM NaH_2PO_4, 1 mM $Na_4P_2O_7$, 0.1 mM ATP, 5 μg of salmon sperm DNA at 42 °C for 20 h. Washing was repeated three times with 2 × SSC, 0.1% SDS for 20 min at 42 °C. Six positive clones were obtained and plaque-purified by secondary screening. Positive clones were rescued according to the manufacturer's instruction and sequenced in both directions.

Site-directed Mutagenesis

The PTP20 mutant containing a cysteine to serine alteration at position 229 was generated using a oligonucleotide primer, CTCTGTGTCCACAGCAGTGCTGGCTGT, according to the protocol of Kunkel (11). The mutation was confirmed by DNA sequencing.

Northern Blot Analysis

Total RNA was extracted from PC12 cells by acid guanidium isothiocyanate-phenol-chloroform method (12). Poly(A)$^+$ RNA was isolated with oligo(dT)-Sepharose (Stratagene) column chromatography according to the manufacturer's instruction. Two micrograms of poly(A)$^+$ RNA was electrophoresed in a formaldehyde, 1.0% agarose gel, blotted to a nitrocellulose membrane filter, and hybridized to ^{32}P-labeled full-length PTP20 cDNA as a probe.

Transient Expression of PTP20 in Rat-1 Cells

The insert of PTP20 was cut out with *Eco*RI digestion and integrated into an expression vector, pcDNA3 (Invitrogen), which had been digested with the same restriction enzyme. The direction of the insert in the plasmid was confirmed by restriction mapping. Rat-1 cells were transfected with the plasmid (2 μg/1 × 10^6 cells) by using Lipofectin (Life Technologies, Inc.). After 48 h of culturing, the cells were washed with phosphate-buffered saline and then lysed with lysis buffer (50 mM HEPES, pH 7.5, containing 150 mM NaCl, 1 mM EDTA, 10% (v/v) glycerol, 1% (v/v) Triton X-100, 1 mM phenylmethylsulfonyl fluoride, 1 mM sodium orthovanadate, 10 μg/ml aprotinin). Protein concentrations of cell lysates were measured with a protein assay kit (Bio-Rad) using bovine serum albumin as a standard. Equivalent amounts of protein were used for Western blot analyses and phosphatase activity assay.

Stable Expression of PTP20 in PC12 Cells

Rat pheochromocytoma cells (PC12) were cultured in Dulbecco's modified Eagle's medium containing high glucose (4.5 g/liter) supplemented with 10% heat-inactivated horse serum and fetal calf serum. 5 × 10^5 cells per 60-mm dish were incubated overnight in 4 ml of growth medium. The following day, the dish was washed once with serum-free medium and then incubated with a Lipofectin (5 μl)-DNA (2 μg) mixture for 6 h. After 48 h, selection started in growth medium containing 500 μg/ml G418 (Life Technologies, Inc.). Following 5 weeks of selection, discrete colonies were subcloned and expanded.

Assay of PTPase Activity

Phosphatase activity of PTP20 was assayed in a 200-μl solution containing 25 mM MES, pH 5.5, 1.6 mM

dithiothreitol, 10 mM *p*-nitrophenyl phosphate (pNPP) as a substrate, and 50 μg of protein cell lysate at 37 °C for 30 min. The reaction was stopped by the addition of 100 μl of 1 N NaOH, and the absorbance was measured at 405 nm.

Western Blot Analyses

Cells were lysed in lysis buffer (detailed above). To assess PTP20 expression, equivalent amounts of protein in the cell lysates were separated by 10% SDS-polyacrylamide gel electrophoresis and electrophoretically transferred to nitrocellulose membranes. The membranes were first incubated with rabbit anti-PTP-PEST antibodies, and then, as secondary antibody, a peroxidase-coupled goat anti-rabbit antibody (Bio-Rad) was added, followed by an enhanced chemiluminescence (ECL) substrate (Amersham) reaction. The substrate reaction was detected on a x-ray film (Amersham). The anti-PTP-PEST antibody was raised against the carboxyl-terminal 56 amino acids of human PTP-PEST (13) which was expressed as a glutathione *S*-transferase fusion protein.

RESULTS AND DISCUSSION

In order to identify novel PTPase genes involved in differentiation of PC12 cells, we performed reversed transcription-based PCR amplification using poly(A)$^+$ RNA from both undifferentiated and differentiated PC12 cells as a template and degenerate primers that were deduced from highly conserved amino acid sequences within the PTPase catalytic domain. This reaction gave specific products of 350-400 bp. These fragments were isolated, cloned, and sequenced. Sequencing analyses revealed the presence of a number of different known PTPase clones including LAR (14), PTPα (LRP) (15), RK-PTP (16), PC12-PTP1 (17), PTP-1B (14), PTP-2E (18), SHP-2 (PTP-1D) (19), rat homologues of SHP-1 (PTP-1C), and MEG2. PCR clone number 20 (PTP20) exhibited sequence similarities but was not identical to any previously known PTPases. Using the PTP20 PCR-generated cDNA fragment as a probe, we isolated a full-length cDNA clone from a PC12 cDNA library and characterized it by sequence analysis.

The cDNA sequence and the deduced amino acid sequence of PTP20 are shown in Fig. 1. The 2226-bp cDNA clone of PTP20 contained an open reading frame of 1359 bp preceded by 27 base pairs of 5′-noncoding region and 840 base pairs of 3′-noncoding region. The 3′-noncoding region contained the polyadenylation signal sequence AATAAA. The open reading frame encoded a protein of 453 amino acids with a predicted molecular mass of approximately 50 kDa. Hydropathy analysis (20) indicated that PTP20 contained no hydrophobic segments appropriate for signal peptide or transmembrane domains, and, therefore, PTP20 is most likely an intracellular protein. The transcripts corresponding to nearly the same size of the full-length cDNA were detected in several rat tissues including brain, liver, lung, spleen, skeletal muscle, kidney, and testis (not shown).

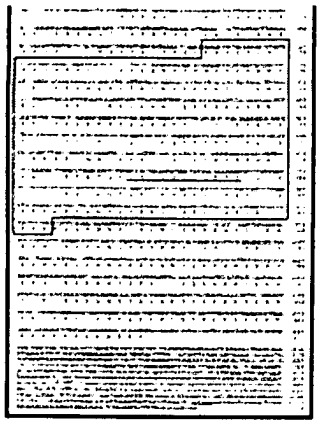


Fig. 1. Nucleotide and deduced amino acid sequences of the cDNA encoding the rat PTP20. Nucleotide and amino acid residues are *numbered* on the *right*. The amino acid consensus sequence (VHC*X*AG*XX*R) found in all PTPases and polyadenylation signal sequences are *underlined*. PTPase domain is *boxed*. An *asterisk* indicates the stop codon.

[View Larger Version of this Image (77K GIF file)]

The PTPase domain was located near the predicted amino terminus between residues 58 and 283. The catalytic domain of PTP20 showed the highest homology with the PTP-PEST family phosphatases, such as human and rat PTP-PESTs (13, 21) and PEP-PTP (22). Proline (P), glutamate (E), serine (S), and threonine (T) residues are enriched in the PEST-motif sequence, which is not arranged in any specified consensus sequence (23). PTP20 may also have a PEST sequence between amino acids 285 and 453, which suggests that PTP20 is a member of the PTP-PEST family. In addition, the last 22 amino acids of PTP20 also had homology to other PTP-PESTs, and, accordingly, antibodies raised against the human PTP-PEST carboxyl terminus cross-react with PTP20. The antibodies were, therefore, used for subsequent experiments as described below.

To confirm that PTP20 encodes a functionally active PTPase, Rat-1 fibroblast cells were transiently transfected with mammalian expression constructs encoding either PTP20 or a Cys → Ser mutant of PTP20. Cell lysates were prepared as described under "Experimental Procedures," and protein concentrations were determined. The expression level of both wild type and catalytically inactive mutant PTP20 was confirmed by Western blotting with anti-PTP-PEST antibodies. Cross-reactivity with nonspecific proteins was not detected as evidenced by lack of signal in the control lane (Fig. 2*A*). Nearly equivalent amounts of expressed protein were detected. The size of the detected protein was 50 kDa which is consistent with the predicted molecular mass of PTP20. For protein-tyrosine phosphatase activity, equivalent amounts of protein from the transfected Rat-1 cell lysates were tested using pNPP as a substrate. Lysates from transfected cells exhibited an approximately 2.5-fold higher PTPase activity over those from control cells, whereas only basal levels of PTPase activity were detected in lysates from cells transfected with a construct encoding a catalytically inactive mutant of PTP20 (Fig. 2*B*). These results indicate that full-length PTP20 cDNA encodes a functionally active PTPase.

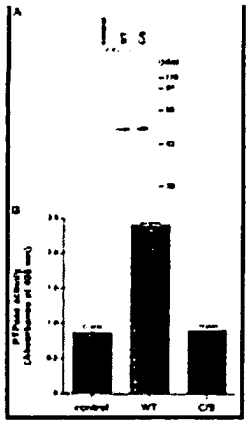


Fig. 2. PTP20 is a functionally active PTPase. *A*, the expression of PC12-PTP (*WT*) and its catalytic mutant (*C/S*) was assessed by Western blotting with antibodies against human PTP-PEST. All lanes contained equivalent amounts of cell lysate (50 μg). *B*, phosphatase activity of PTP20 was determined by hydrolysis of the PTPase substrate, pNPP. Hydrolysis of pNPP was quantified by measuring the absorbance at 405 nm. Cell lysates were prepared from Rat-1 cells which were transiently transfected with the expression vector pcDNA3 alone (control), the PTP20 expression vector (*WT*), or with an expression vector encoding a catalytic mutant of PTP20 (*C/S*). The data are shown as the average ± S.E. of three independent experiments.

[View Larger Version of this Image (26K GIF file)]

To elucidate the role of PTP20 in the differentiation process of PC12 cells, Northern blot analysis was used to examine the expression pattern of PTP20 mRNA in PC12 cells treated with NGF for 3 or 6 days. Untreated PC12 cells exhibited a 2.3-kb PTP20 mRNA transcript (Fig. 3). Following 3 days of NGF treatment, a 1.5-fold increase in the amount of transcript was observed. Another 3 days of NGF treatment caused a 2.4-fold increase as compared to untreated cells. In addition to the predominant 2.3-kb transcript, a faint band of 1.5 kb in size was also detected which also increased in abundance as NGF treatment continued. The expression pattern of PTP20 mRNA suggested that PTP20 might play a role during NGF-induced PC12 differentiation.

Fig. 3. Expression of PTP20 in PC12 cells. *A*, Northern blot analysis was curried out with 2 μg of poly(A)$^+$ RNA from PC12 cells treated with NGF (50 ng/ml) for 0, 3, or 6 days. *B*, the expression level of the 2.3-kb transcript was quantified with a PhosphorImager (Fuji Film). The data are representative of three independent experiments.

[View Larger Version of this Image (28K GIF file)]

To further elucidate the function of PTP20 in the differentiation process, PC12 cells were stably transfected with the PTP20 cDNA mammalian expression construct. From 7 stably transfected clonal cell lines, three independent clones showed high levels of PTP20 expression as assessed by Western blotting (Fig. 4*A*, *lanes 2-4*). In parental PC12 cells, endogenous PTP20 protein was beneath detection with the antibody (Fig. 4*A*, *lane 1*). These three independent clones appeared morphologically indistinguishable from parental PC12 cells (Fig. 4*B*). However, following NGF treatment (50 ng/ml), all three clones showed accelerated neurite outgrowth, with 20 to 40% of the cells expressing neurites of more than two cell bodies in length at day 1 and more than 70% of the cells expressing such neurites at day 3. In contrast, the parental PC12 cells

showed less than 5% of the cells with neurites of two cell bodies in length at day 1 and 47% at day 3. At day 4 following NGF treatment, more than 70% of both parental PC12 cells and PTP-PC12 cells expressed neurite outgrowth; however, the neurite length and the abundance of neurites in PTP-PC12 cells appeared longer and larger than those of parental PC12 cells.

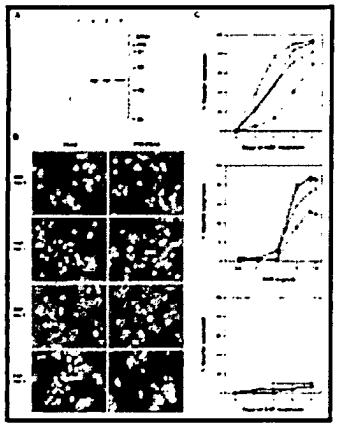


Fig. 4. Effect of overexpression of PTP20 on the differentiation of PC12 cells induced by NGF. *A*, expression of PTP20 in PC12 cells. PC12 cells were stably transfected with PTP20 expression plasmid. Following selection in medium supplemented with G418, surviving colonies were subcultured and the level of PTP20 was assessed by Western blotting with anti-human PTP-PEST antibodies. The transfected cells were named PTP-PC12, and three independent clones (PC-YN1, -YN2, and -YN3) were tested. All lanes contained equivalent amounts of cell lysate (50 μg). *Lane 1*, PC12; *2*, PC-YN1; *3*, PC-YN2; *4*, PC-YN3. *B*, effects of NGF or EGF treatment on PC12 and PTP-PC12 cells (PC-YN1). Cells were cultured in the medium supplemented with either NGF (50 ng/ml) or EGF (50 ng/ml), and photographs were taken on the day as indicated. Magnification × 100. *C*, response of PC12 and PTP-PC12 cells to NGF and EGF treatment. At the indicated days, 100 to 200 cells were counted and scored for the expression of neurites longer than one cell body. •, PC12; ○, PC-YN1; △, PC-YN2; □, PC-YN3. *Upper panel*, kinetics of NGF-induced neurite outgrowth in response to NGF (50 ng/ml); *middle panel*, dose-dependent response to NGF. Cells were counted after 3 days of treatment; *lower panel*, kinetics of EGF-induced neurite outgrowth in response to EGF (50 ng/ml). All data shown here are representative of three independent experiments.
[View Larger Version of this Image (72K GIF file)]

It has been reported that epidermal growth factor (EGF) triggers neuronal differentiation of PC12 cells when EGF receptors were overexpressed (24). Hanafusa and co-workers (25) have reported that, when overexpressed in PC12 cells, v-Crk accelerates differentiation induced not only by NGF but also by EGF. To test the effect of EGF on PTP-PC12 cells, the cells were treated with EGF (50 ng/ml). This concentration of EGF was chosen because it was consistent with the amounts of EGF used in experiments by Hanafusa and co-workers (25). Even after 4 days of EGF treatment, fewer than 5% of the PTP-PC12 cells and the parental PC12 cells exhibited any morphological changes (Fig. 4, *B* and *C*, respectively). It is therefore unlikely that EGF stimulates neurite outgrowth of PTP-PC12 cells, in contrast to the previous observations in PC12 cells expressing v-Crk or the EGF receptor.

EGF- and NGF-treated PC12 cells share several common properties including rapid membrane ruffling, cell flattening, and increases cell adhesion (26). However, only NGF promotes neurite outgrowth. This difference was originally thought to be mediated by distinct intracellular signaling pathways stimulated by NGF (27). Recent studies have demonstrated, however, that the differential effects of NGF and EGF on PC12 cell differentiation may be defined by differences in the extent and duration of activation of the same pathway leading to MAP kinase activation (24, 28, 29). This strongly suggested the critical involvement of a phosphatase in this process. In this study, we demonstrated that overexpression of the protein-tyrosine phosphatase PTP20 causes accelerated differentiation of PC12 cells following NGF stimulation. These data

suggest an involvement of PTP20 in the regulation of the signal transduction pathway leading to neuronal differentiation upon NGF stimulation. The target(s) of the action of PTP20 may therefore represent one or more components of the MAP kinase cascade which upon dephosphorylation remove a negative constraint and thereby result in a sustained MAP kinase signal. Alternatively, the role of PTP20 may be to negatively control cell proliferation and thereby promote differentiation, since preliminary data indicated that PTP-PC12 cells proliferate more slowly than parental PC12 cells in the absence of exogenous growth factor addition (data not shown).

It has been reported that the expression level of some PTPases such as LAR (28), PTP-P1 (29), and PC12-PTP1 (17) is up-regulated during NGF-induced differentiation of PC12 cells. The results presented here indicate similar findings for PTP20 and furthermore demonstrated that ectopic overexpression of this PTPase promotes the differentiation process in PC12 cells. This is the first demonstration of a specific PTPase to be involved in neuronal differentiation of PC12 cells. Whether these findings represent a specific characteristic of PC12 cells or reflect an analogous role of PTP20 in the development and maintenance of the nervous system remains to be established.

FOOTNOTES

* This work was supported by a grant from Sugen Inc. (to A. U.). The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) U69673[GenBank].

‡ These authors contributed equally to this work.
§ To whom correspondence should be addressed. Tel: 49-89-8578-2513; Fax: 49-89-857-7866.

1 The abbreviations used are: PTPase(s), protein-tyrosine phosphatase(s); NGF, nerve growth factor; pNPP, *para*-nitrophenyl phosphate; PCR, polymerase chain reaction; MAP, mitogen-activated protein; EGF, epidermal growth factor; MES, 2-(*N*-morpholino)ethanesulfonic acid; bp, base pair(s); kb, kilobase(s).

Acknowledgments

We thank Dr. Irmi Sures for providing degenerate primers and Dr. Karen J. Martell for valuable editorial comments on this manuscript.

REFERENCES

1. Ullrich, A., and Schlessinger, J. (1990) *Cell* **61**, 203-212 [Medline]

2. Yarden, Y., and Ullrich, A. (1988) *Annu. Rev. Biochem.* **57**, 443-478 [Medline]
3. Walton, K. M., and Dixon, J. E. (1993) *Annu. Rev. Biochem.* **62**, 101-120 [Medline]
4. Klein, R., Jing, S., Nabduri, V., O'Rourke, E., and Barbacid, M. (1991) *Cell* **65**, 189-197 [Medline]
5. Kaplan, D. R., Hempstead, B. L., Martin-Zanca, D., Chao, M. V., and Parada, L. F. (1991) *Science* **252**, 554-558 [Medline]
6. Nabreda, A. R., Martin-Zanca, D., Kaplan, D. R., Parada, L. F., and Santos, E. (1991) *Science* **252**, 558-561 [Medline]
7. Greene, L. A., and Tischler, A. S. (1982) *Adv. Cell Neurobiol.* **3**, 373-415
8. Wood, K. W., Sarnecki, C., Roberts, T. M., and Blenis, J. (1992) *Cell* **68**, 1041-1050 [Medline]
9. Thomas, S. M., Demarco, M., D'Arcangelo, G., Hakgoua, S., and Bruhhe, J. S. (1992) *Cell* **68**, 1031-1040 [Medline]
10. Wu, Y. Y., and Bradshaw, R. A. (1993) *J. Cell Biol.* **121**, 409-422 [Abstract]
11. Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 488-492 [Medline]
12. Chomczynski, P., and Sacci, N. (1987) *Anal. Biochem.* **162**, 156-159 [CrossRef][Medline]
13. Takekawa, M., Itoh, F., Hinoda, Y., Arimura, Y., Toyota, M., Sekiya, M., Adachi, M., Imai, K., and Yachi, A. (1992) *Biochem. Biophys. Res. Commun.* **189**, 1223-1230 [Medline]
14. Hashimoto, N., Zhang, W. R., and Goldstein, B. J. (1992) *Biochem. J.* **284**, 569-576 [Medline]
15. Moriyama, T., Fujiwara, Y., Imai, E., Takenaka, M., Kawanishi, S., Inoue, T., Noguchi, T., Tanaka, T., Kamada, T., and Ueda, N. (1992) *Biochem. Biophys. Res. Commun.* **188**, 34-39 [Medline]
16. Moriyama, T., Kawanishi, S., Inoue, T., Imai, E., Kaneko, T., Xia, C., Takenaka, M., Noguchi, T., Kamada, T., and Ueda, N. (1994) *FEBS Lett.* **353**, 305-308 [Medline]
17. Sharma, E., and Lombroso, P. J. (1995) *J. Biol. Chem.* **270**, 49-53 [Abstract/Full Text]
18. L'Abbe, D., Banville, D., Tong, Y., Stocco, R., Masson, S., Ma, S., Fantus, G., and Shen, S. H. (1994) *FEBS Lett.* **356**, 351-356 [Medline]
19. Mei, L., Dorherty, C. A., and Huganir, R. L. (1994) *J. Biol. Chem.* **269**, 12254-12262 [Abstract]
20. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105-132 [Medline]
21. Yang, Q., Co, D., Sommercorn, J., and Tonks, N. K. (1993) *J. Biol. Chem.* **268**, 6622-6628 [Abstract]
22. Matthews, R. J., Bowne, D. B., Flores, E., and Thomas, M. L. (1992) *Mol. Cell. Biol.* **12**, 2396-2405 [Abstract]
23. Rechsteiner, M., and Rogers, S. W. (1996) *Trends Biochem. Sci.* **21**, 267-271 [CrossRef][Medline]
24. Traverse, S., Seedorf, K., Paterson, H., Marshall, C. J., Cohen, P., and Ullrich, A. (1994) *Curr. Biol.* **4**, 694-701 [Medline]
25. Hempstead, B. L., Birge, R. B., Fajardo, J. E., Glassman, R., Mahadeo, D., Kraemer, R., and Hanafusa, H. (1994) *Mol. Cell. Biol.* **14**, 1964-1971 [Abstract]
26. Connolly, J. L., Greene, S. A., and Greene, L. A. (1984) *J. Cell Biol.* **98**, 457-465 [Abstract]
27. Chao, M. V. (1992) *Cell* **68**, 995-997 [Medline]
28. Dikic, I., Schlessinger, J., and Lax, I. (1994) *Curr. Biol.* **4**, 702-708 [Medline]
29. Obermeier, A., Bradshaw, R. A., Seedorf, K., Choidas, A., Schlessinger, J., and Ullrich, A. (1994) *EMBO J.* **13**, 1585-1590 [Abstract]
30. Longo, F. M., Martignetti, J. A., LeBeau, J. M., Zhang, J. S., Barnes, J. P., and Brosius, J. (1993) *J. Biol. Chem.* **268**, 26503-26511 [Abstract]
31. Pan, M.-G., Rim, C., Lu, K. P., Florio, T., and Stork, P. J. S. (1993) *J. Biol. Chem.* **268**, 19284-19291 [Abstract]

This article has been cited by other articles:

- Abstract of this Article
- **Reprint (PDF) Version of this Article**
- Similar articles found in:
  - JBC Online
  - PubMed
- PubMed Citation
- This Article has been cited by:
- Search Medline for articles by:
  - Aoki, N. || Ullrich, A.
- Alert me when:
  - new articles cite this article
- Download to Citation Manager

- Spencer, S., Dowbenko, D., Cheng, J., Li, W., Brush, J., Utzig, S., Simanis, V., Lasky, L. A. (1997). PSTPIP: A Tyrosine Phosphorylated Cleavage Furrow-associated Protein that Is a Substrate for a PEST Tyrosine Phosphatase. *J. Cell Biol.* 138: 845-860 [Abstract] [Full Text]
- Lei, S., Dryden, W. F., Smith, P. A. (1997). Regulation of N- and L-Type Ca2+ Channels in Adult Frog Sympathetic Ganglion B Cells by Nerve Growth Factor In Vitro and In Vivo. *J. Neurophysiol.* 78: 3359-3370 [Abstract] [Full Text]
- Wang, B., Lemay, S., Tsai, S., Veillette, A. (2001). SH2 Domain-Mediated Interaction of Inhibitory Protein Tyrosine Kinase Csk with Protein Tyrosine Phosphatase-HSCF. *Mol. Cell. Biol.* 21: 1077-1088 [Abstract] [Full Text]
- Davidson, D., Cloutier, J.-F., Gregorieff, A., Veillette, A. (1997). Inhibitory Tyrosine Protein Kinase p50csk Is Associated with Protein-tyrosine Phosphatase PTP-PEST in Hemopoietic and Non-hemopoietic Cells. *J. Biol. Chem.* 272: 23455-23462 [Abstract] [Full Text]
- Ihara, S., Nakajima, K., Fukada, T., Hibi, M., Nagata, S., Hirano, T., Fukui, Y. (1997). Dual control of neurite outgrowth by STAT3 and MAP kinase in PC12 cells stimulated with interleukin-6. *EMBO J.* 16: 5345-5352 [Abstract] [Full Text]
- Aoki, N., Matsuda, T. (2000). A Cytosolic Protein-tyrosine Phosphatase PTP1B Specifically Dephosphorylates and Deactivates Prolactin-activated STAT5a and STAT5b. *J. Biol. Chem.* 275: 39718-39726 [Abstract] [Full Text]
- Aoyama, K., Nagata, M., Oshima, K., Matsuda, T., Aoki, N. (2001). Molecular Cloning and Characterization of a Novel Dual Specificity Phosphatase, LMW-DSP2, That Lacks the Cdc25 Homology Domain. *J. Biol. Chem.* 276: 27575-27583 [Abstract] [Full Text]

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 276, No. 28, Issue of July 13, pp. 26622–26628, 2001
Printed in U.S.A.

The Ras/Rac Guanine Nucleotide Exchange Factor Mammalian Son-of-sevenless Interacts with PACSIN 1/Syndapin I, a Regulator of Endocytosis and the Actin Cytoskeleton*

Received for publication, January 22, 2001, and in revised form, April 30, 2001
Published, JBC Papers in Press, May 14, 2001, DOI 10.1074/jbc.M100591200

Sylwia Wasiak‡§, Christopher C. Quinn¶‖, Brigitte Ritter, Elaine de Heuvel‡, Danny Baranes‡‡§§, Markus Plomann**, and Peter S. McPherson‡ ‡‡ ¶¶**

*From the ‡Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal H3A 2B4, Quebec, Canada, the ¶Section of Neurobiology, Yale University School of Medicine, New Haven, Connecticut 06510, the **Institute for Biochemistry, University of Cologne, D-50931 Cologne, Germany, and the ‡‡Department of Anatomy and Cell Biology, McGill University, Montreal H3A 2B2, Quebec, Canada*

Mammalian Son-of-sevenless (mSos) functions as a guanine nucleotide exchange factor for Ras and Rac, thus regulating signaling to mitogen-activated protein kinases and actin dynamics. In the current study, we have identified a new mSos-binding protein of 50 kDa (p50) that interacts with the mSos1 proline-rich domain. Mass spectrometry analysis and immunodepletion studies reveal p50 as PACSIN 1/syndapin I, a Src homology 3 domain-containing protein functioning in endocytosis and regulation of actin dynamics. In addition to PACSIN 1, which is neuron-specific, mSos also interacts with PACSIN 2, which is expressed in neuronal and nonneuronal tissues. PACSIN 2 shows enhanced binding to the mSos proline-rich domain in pull-down assays from brain extracts as compared with lung extracts, suggesting a tissue-specific regulation of the interaction. Proline to leucine mutations within the Src homology 3 domains of PACSIN 1 and 2 abolish their binding to mSos, demonstrating the specificity of the interactions. *In situ*, PACSIN 1 and mSos1 are co-expressed in growth cones and actin-rich filopodia in hippocampal and dorsal root ganglion neurons, and the two proteins co-immunoprecipitate from brain extracts. Moreover, epidermal growth factor treatment of COS-7 cells causes co-localization of PACSIN 1 and mSos1 in actin-rich membrane ruffles, and their interaction is regulated through epidermal growth factor-stimulated mSos1 phosphorylation. These data suggest that PACSINs may function with mSos1 in regulation of actin dynamics.

Ras functions as a molecular switch in the transduction of a wide variety of growth and differentiation signals induced by extracellular ligands (1). Activation of Ras is mediated by several Ras-specific guanine nucleotide exchange factors (GEFs)[1] that convert GDP-Ras into GTP-Ras (1). Prominent among these is mammalian Son-of-sevenless (mSos). mSos interacts through a C-terminal proline-rich domain (PRD) with the Src homology 3 (SH3) domains of Grb2, an adaptor protein that targets mSos to activated growth factor receptors (2–7). Additionally, mSos interacts through the PRD with the endocytic adaptor proteins amphiphysin II (8) and intersectin (9, 10). These interactions may function to target mSos to Ras activation on the endocytic pathway (11–13).

In addition to the PRD, mSos contains a CDC25 homology domain, encoding Ras GEF activity (3–5, 14), and a Dbl homology domain, endowed with GEF activity for Rac, a member of the Rho superfamily of GTPases (15, 16). In fact, mSos is the prototype member of a family of bifunctional GEFs, including Ras-GRF1 and Ras GRF-2, having dual specificity for Ras and Rac (16). Through its Dbl homology domain, mSos binds directly to Rac (17). However, the GEF activity of mSos toward Rac appears to be unique relative to other Rho family GEFs in that it catalyzes guanine nucleotide exchange as part of a macromolecular complex with Eps8 and E3b1, two proteins functioning in growth factor signaling (18, 19). Rac activation has multiple effects in cells, the most prominent being alterations in the actin cytoskeleton leading to membrane ruffling and lamellipodia formation (20). Thus mSos, through its ability to activate Rac, is thought to play a functional role in growth factor-mediated regulation of actin dynamics (18, 21).

To screen for novel mSos binding partners, we performed overlay assays of brain extracts with fusion proteins encoding the PRD of mSos1. A major mSos1-binding protein of 50 kDa was detected and purified by affinity chromatography. Mass spectrometry analysis identified the protein as PACSIN 1/syndapin I. PACSIN 1 was originally identified based on its differential expression in intact and lesioned mouse brain (22), and syndapin I was independently identified through its SH3 domain-dependent interaction with dynamin 1 (23). We will

* This work was supported by Canadian Institutes of Health Research Operating Grant MT-15396 (to P. S. M.) and the Köln Fortune Program of the Medical Faculty of the University of Cologne (to M. P.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Supported by studentships from the Natural Sciences and Engineering Research Council and from Fonds de Recherche en Sante du Quebec.

‖ Supported by National Institutes of Health, NINDS Grant NS22807 (to Dr. Susan Hockfield of Yale University).

§§ Present address: Dept. of Life Sciences, Ben-Gurion University, Beer Sheva 84105, Israel.

¶¶ Supported by an Investigator Award from the Canadian Institutes of Health Research. To whom correspondence should be addressed: Dept. of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, 3801 rue University, Montreal H3A 2B4, Quebec, Canada. Tel.: 514-398-7355; Fax: 514-398-8106; E-mail: mcpm@musica.mcgill.ca.

[1] The abbreviations used are: GEF, guanine nucleotide exchange factor; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonic acid; CT, C-terminal; EGF, epidermal growth factor; GST, glutathione *S*-transferase; MEK, mitogen-activated protein kinase/extracellular signal-regulated kinase kinase; mSos, mammalian son-of-sevenless; NT, N-terminal; PAGE, polyacrylamide gel electrophoresis; PBS, phosphate-buffered saline; PRD, proline-rich domain; SH, Src homology.

EXHIBIT F

use the name PACSIN throughout to collectively refer to PACSIN and syndapin. Whereas PACSIN 1 is neuron-specific, its closely related homologue PACSIN 2 is expressed in brain and several nonneuronal tissues (24, 25). Interestingly, the PACSINs appear to be involved in regulation of endocytosis and the actin cytoskeleton. Through a C-terminal SH3 domain, the PACSIN isoforms interact with the endocytic regulatory enzymes dynamin 1 and synaptojanin 1, as well as with N-WASP, a stimulator of Arp2/3-mediated actin nucleation and assembly (23, 25, 26). Overexpression of full-length PACSIN stimulates cortical actin assembly, leading to filopodia formation, and the PACSINs localize to sites of high actin turnover, such as filopodia and lamellipodia (25).

After identifying PACSIN 1 as a mSos1 binding partner, we confirmed the interaction *in vitro* and used co-immunoprecipitation analysis to demonstrate the interaction *in vivo*. Interestingly, mSos1 co-distributes with PACSIN 1 in the growth cones and filopodia of cultured hippocampal neurons, and both proteins co-localize with actin in the filopodia from growth cones of dorsal root ganglia neurons in culture. Further, PACSIN 1 and mSos1 are co-localized in growth factor-induced membrane ruffles in COS-7 cells, and their interaction is regulated by mSos1 phosphorylation. Together, these data provide further evidence for a role for mSos in regulation of the actin cytoskeleton.

EXPERIMENTAL PROCEDURES

Antibodies—Affinity-purified antibodies against PACSIN 2 (26) and amphiphysin I and II (27) were described previously. Polyclonal antiserum 2704 against rat syndapin I was a generous gift of Dr. Regis Kelly (University of California, San Francisco) (23). Monoclonal antibodies against α-tubulin and FLAG epitope (Sigma), tetra-His epitope (Qiagen), dynamin 1 (HUDY-1) (Upstate Biotechnology Inc.), and a polyclonal antibody C23 against mSos1 (Santa Cruz Biotechnology) were obtained commercially.

DNA Constructs and Recombinant Proteins—His_6-tagged rat syndapin I (25) and Flag-tagged mouse Sos1 (28) in mammalian expression vectors were generous gifts of Dr. Regis Kelly (University of California, San Francisco) and Dr. Jeffrey Pessin (University of Iowa), respectively. A protein construct encoding the SH3 domain of rat syndapin I (residues 376–441) was generated by polymerase chain reaction with Vent DNA polymerase (New England Biolabs) using full-length cDNA as template and the following primers: forward, 5′-CGCCTCGAGCGGATCCAACCCCTTCGAGGACGATGC-3′; reverse, 5′-CGGAATTCCTATATAGCCTCAACGTAGTTG-3′. The resulting polymerase chain reaction product was digested with *Bam*HI and *Eco*RI and cloned in-frame into the corresponding sites of pGEX-2T. Mouse Sos1 cDNA was used as a template to generate the following GST fusion proteins: GST-NT (residues 1111–1228) and GST-CT (residues 1223–1341). GST-NT was generated with the forward primer 5′-GCGGATCCTCTGGCACCTCCAGCAAC-3′ and the reverse primer 5′-GCGGAATTCTCAATCAGGTGTCCTCACAGG-3′. GST-CT was generated with the forward primer 5′-GCGGGATCCCCTGTGAGGACACCTGATG-3′ and the reverse primer 5′-GCGGAATTCTCAGGAAGAATGGGCATTC-3′. The resulting polymerase chain reaction products were digested with *Bam*HI and *Eco*RI and cloned in-frame into the corresponding sites of pGEX-2T. GST fusion proteins encoding full-length mouse PACSIN 1 and 2 were prepared as described (26). PACSIN isoforms containing single amino acid changes in the SH3 domains were derived using the mutation oligonucleotides P1-P434L (5′-GGCCTCTATCTCGCGAACTACGTTG-3′) for PACSIN 1 and P2-P478L (5′-GGCCTATACCTCGCGAACTATGTCG-3′) for PACSIN 2 on the corresponding wild-type cDNAs in combination with the Transformer™ site-directed mutagenesis kit (CLONTECH).

Tissue and Subcellular Fractionation—Various adult rat tissues, including brain, were homogenized in Buffer A (10 mM HEPES-OH, pH 7.4, 0.83 mM benzamidine, 0.23 mM phenylmethylsulfonyl fluoride, 0.5 μg/ml aprotinin, and 0.5 μg/ml leupeptin). A postnuclear supernatant was obtained by centrifugation for 5 min at 800 × g_{max}, and the extracts were then separated into cytosolic and membrane fractions by ultracentrifugation at 205 000 × g_{max} for 30 min at 4 °C. In some cases, the postnuclear supernatant was incubated with 1% Triton X-100 for 30 min prior to ultracentrifugation, leading to a crude Triton-soluble lysate. Differential centrifugation of rat brain extracts, leading to the defined subcellular fractions in Fig. 2, was performed as described previously (29).

Overlay Assays—Overlay assays with GST fusion proteins were performed as described (30). Briefly, protein fractions were resolved by SDS-PAGE and transferred to nitrocellulose. Membranes were blocked in 5% nonfat dry milk powder in phosphate-buffered saline (PBS) (20 mM NaH_2PO_4, 0.9% NaCl, pH 7.4) for 1 h and incubated overnight at 4 °C with 10 or 20 μg of GST fusion protein diluted in Tris-buffered saline (20 mM Tris-Cl, 150 mM NaCl, pH 7.4) containing 3% bovine serum albumin, 0.1% Tween-20, and 1 mM dithiothreitol. Bound fusion protein was subsequently detected using affinity-purified antibodies directed against GST.

Affinity Chromatography—Cytosolic or crude Triton-soluble lysates were prepared from different tissues as described above. For the cytosolic samples, Triton X-100 was added to 1%. The samples were then incubated for 2 h or overnight at 4 °C with GST fusion proteins precoupled to glutathione-Sepharose beads. After incubation, samples were washed three times in Buffer A containing 1% Triton X-100, and bound proteins were resolved by SDS-PAGE and processed for Western blot analysis or stained with Coomassie Blue. For precipitation experiments with full-length fusion proteins of wild-type and mutant PACSINs, mouse brains were homogenized in Buffer B (10 mM HEPES, pH 7.4, 150 mM NaCl, 1 mM EGTA, 0.1 mM $MgCl_2$,) containing 1% CHAPS and a protease inhibitor mixture (Sigma). The homogenates were centrifuged for 30 min at 21,000 × g, the supernatant was decanted and recentrifuged, and Triton X-100 was added to the resulting supernatant at a final concentration of 0.05%. The preparation was dialyzed overnight against Buffer B and centrifuged as before. The resulting supernatant was incubated overnight at 4 °C with GST-PACSINs precoupled to glutathione-Sepharose. The beads were subsequently washed extensively in Buffer B containing 0.1% Triton X-100, and bound proteins were resolved by SDS-PAGE and processed for Western blot analysis. For phosphorylation experiments, COS-7 cells were serum-starved overnight and preincubated with 50 μM PD-098059 (Calbiochem, La Jolla, CA) in Me_2SO or Me_2SO alone for 30 min at 37 °C. Cells were then treated with 100 ng/ml epidermal growth factor (EGF) for 5 min at 37 °C in the absence or presence of PD-098059 and lysed in ice-cold Buffer C (10 mM HEPES-OH, pH 7.4, 5 mM EGTA, 5 mM EDTA, 50 mM sodium fluoride, 20 mM sodium pyrophosphate, 1 mM sodium vanadate, 0.83 mM benzamidine, 0.23 mM phenylmethylsulfonyl fluoride, 0.5 μg/ml aprotinin, and 0.5 μg/ml leupeptin). Cell lysates were solubilized with 1% Triton X-100 and centrifuged in a Beckman TLA 100.2 rotor at 75,000 rpm for 15 min. Cell extracts were incubated for 1 h at 4 °C with GST fusion proteins (~5 μg/ml of each protein) prebound to glutathione-Sepharose beads. After incubation, samples were washed three times in Buffer C containing 1% Triton X-100, and bound proteins were resolved by SDS-PAGE and processed for Western blot analysis.

Immunoprecipitation Analysis—Cytosolic or crude Triton-soluble lysates were prepared from different tissues as described above. For the cytosolic samples, Triton X-100 was added to 1%. The sample were then precleared by incubation with protein A-Sepharose, and an aliquot of the precleared extract was incubated overnight at 4 °C with normal rabbit serum or different antibodies precoupled to protein A-Sepharose. Beads were washed in Buffer A containing 1% Triton X-100, and proteins specifically bound to the beads were eluted and processed for SDS-PAGE. For immunodepletion experiments, an identical protocol was used except that 100 μg of cytosolic extract was added to the beads and the material that did not bind to the beads was processed for SDS-PAGE.

Primary Cell Culture—Dissociated cell cultures were prepared from the CA3 and dentate regions of hippocampi from P1 rat pups as described (31). Following 2 days in culture, the hippocampal neurons were fixed with PBS containing 4% paraformaldehyde and 4% sucrose for 15 min at room temperature. The cells were permeabilized with 0.1% Triton X-100 for 5 min and blocked with PBS containing 1% normal goat serum. For explant cultures, dorsal root ganglia were dissected from E15 Harlan Sprague-Dawley rats. The ganglia were cut in half and placed on coverslips coated with 10 μg/ml natural mouse laminin (Life Technologies, Inc.). The cultures were incubated in F-12 culture medium containing 10% fetal calf serum supplemented with 50 ng/ml 7S nerve growth factor. Following 1–2 days in culture, the cells were fixed with PBS containing 3% paraformaldehyde and 0.3 M sucrose for 30 min at room temperature. The cells were permeabilized with 0.2% Triton X-100 for 3 min and blocked with PBS containing 5% bovine serum albumin and 5% normal goat serum. Following blocking, both culture types were processed for immunofluorescence with antibodies 2704 and C23. In some cases, filamentous actin was detected with

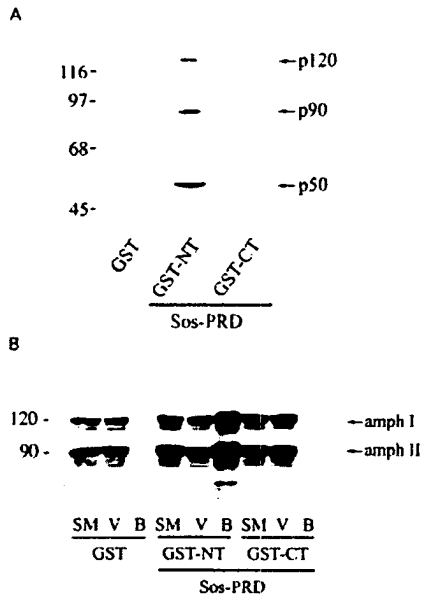


FIG. 1. **Overlay analysis of mSos1-binding proteins.** *A*, proteins of cytosolic fractions from rat brain were separated by SDS-PAGE, transferred to nitrocellulose, and overlaid with GST fused to amino acids 1111–1228 (*GST-NT*) or amino acids 1223–1341 (*GST-CT*) of the PRD of mouse Sos1 or with GST alone. The migratory positions of three major bands that bind to GST-NT and that of the molecular weight standards are indicated on the *right* and *left*, respectively. *B*, a crude Triton X-100-soluble rat brain extract was incubated with GST, GST-NT, or GST-CT conjugated to glutathione-Sepharose beads. Proteins specifically bound to the beads (*B*) along with aliquots of the brain extract (starting material (*SM*)) and equal amounts of the unbound material (void (*V*)) were processed for Western blot with an antibody that recognizes both amphiphysin I (*amph I*) and amphiphysin II (*amph II*). The molecular masses of amphiphysin I and II are indicated on the *left*.

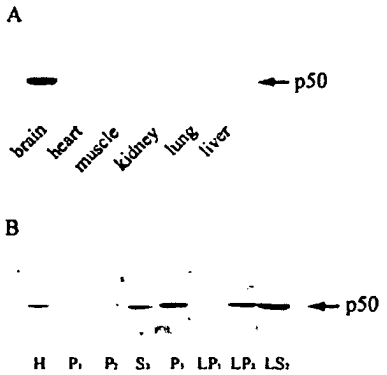


FIG. 2. **Tissue and subcellular distribution of p50.** *A*, proteins of crude Triton X-100-soluble extracts from various adult rat tissues (200 μg/tissue) were separated by SDS-PAGE, transferred to nitrocellulose, and overlaid with GST-NT. *B*, proteins of brain subcellular fractions (100 μg/fraction) were separated by SDS-PAGE, transferred to nitrocellulose, and overlaid with GST-NT. Subcellular fractions were prepared as described (28). *H*, homogenate; *P*, pellet; *S*, supernatant; *LP*, lysed pellet; *LS*, lysed supernatant. For both *A* and *B*, the migratory position of p50 is indicated by the *arrow* on the *right*.

phalloidin-Alexa 488 (Jackson ImmunoResearch).

Immunofluorescence on COS-7 Cells—COS-7 cells were plated on poly-L-lysine-coated coverslips, transfected with LipofectAMINE 2000 (Life Technologies, Inc.), serum-starved overnight, and then left untreated or stimulated with 100 ng/ml EGF for 2 min at 37 °C. Cells were washed twice in ice-cold PBS and processed for immunofluorescence as previously described (32) using antibodies 2704 or C23. Filamentous actin was detected with phalloidin-tetramethylrhodamine isothiocyanate (Sigma).

RESULTS AND DISCUSSION

We previously demonstrated that mSos interacts through its PRD with the endocytic protein intersectin, suggesting that intersectin may target mSos to Ras on the endocytic pathway (9, 10, 13). As the PRD of mSos contains multiple SH3 domain-binding consensus sites, we sought to identify additional mSos binding partners. Overlay of adult rat brain extracts with a GST fusion protein encoding the N-terminal half of the mouse Sos1 PRD (GST-NT) (amino acids 1111–1228) identified three proteins of 120 (p120), 90 (p90), and 50 (p50) kDa (Fig. 1*A*). None of the bands were detected with a GST fusion protein encoding the mouse Sos1 PRD C-terminal half (GST-CT) (amino acids 1223–1341) or with GST alone (Fig. 1*A*). Previously, Leprince *et al.* (8) identified amphiphysin II as a mSos1 binding partner. To determine whether p120 and p90 correspond to amphiphysin I (120 kDa) and amphiphysin II (90 kDa), respectively, we used GST-NT, GST-CT, or GST alone in pull-down assays with soluble rat brain extracts. Western blots of the pull-downs demonstrated that both amphiphysin I and II bind specifically to GST-NT (Fig. 1*B*), suggesting that they represent p120 and p90. Intersectin also bound selectively to GST-NT, whereas Grb2 bound equally well to GST-NT and GST-CT (data not shown). Surprisingly, neither Grb2 nor intersectin was detected on the overlay assays, possibly due to lower levels of expression in brain extracts than the amphiphysins or p50. The abundant SH3 domain-containing protein, endophilin 1, which is readily detected on overlays with the PRDs of synaptojanin 1 (33) and dynamin 1 (34), was not seen on the overlays with the PRD of mSos1, further demonstrating the specificity of the interactions detected.

To characterize p50, the major mSos1 binding partner identified, we performed overlays with GST-NT on tissue extracts. p50 was detected in brain but not in a variety of nonneuronal tissues (Fig. 2*A*). This is consistent with the distribution of mSos1 that is expressed at higher levels in brain than in other tissues (9). Within brain, subcellular fractionation revealed p50 in both soluble and particulate fractions (Fig. 2*B*). The greatest enrichment was seen in the second lysed supernatant fraction (Fig. 2*B*, LS_2), which contained soluble proteins generated from the lysis of crude synaptosomes. This distribution is similar to that previously described for mSos1 (9), as well as that of the presynaptically enriched endocytic regulatory enzymes dynamin 1 and synaptojanin 1 (29).

To identify p50, we used the mSos1 GST fusion proteins to affinity purify mSos1-binding proteins from a soluble rat brain extract. As determined by Coomassie Blue staining, a 50-kDa band that bound to GST-NT but not to GST-CT or to GST alone was the major affinity-selected protein (Fig. 3). Minor bands at 120, 90, and 70 kDa were also weakly detected. The 50-kDa band was excised from the gel and subjected to trypsin digestion, and the fragments were analyzed by matrix assisted laser desorption ionization mass spectrometry at the W. M. Keck Foundation Biotechnology Resource Laboratory at Yale University. A ProFound search of the peptide masses provided a tentative identification for p50 as PACSIN 1/syndapin I. PACSIN 1 was identified based on its up-regulation during neuronal differentiation in mouse (22), whereas its rat orthologue syndapin I was identified through its SH3 domain-dependent interaction with dynamin 1 (23). PACSIN 1, which contains an SH3 domain at its C terminus, is a neuron-specific protein with a predicted molecular mass of 50 kDa, consistent with its identification as p50. To support this identification, we performed overlay assays with GST-NT or GST alone on cells transfected with a cDNA encoding His_6-tagged full-length PACSIN 1. GST-NT specifically interacted with a protein species that perfectly co-migrated with PACSIN 1, as detected with an anti-His_6 Western blot (Fig. 4*A*), demonstrating that

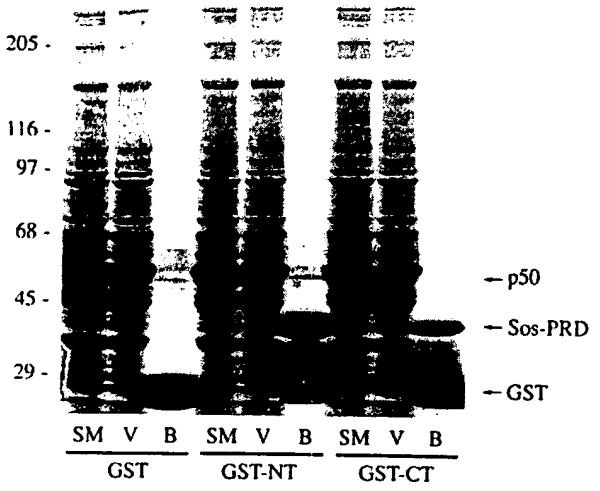


FIG. 3. **Affinity purification of p50.** A rat brain cytosolic extract was incubated with GST, GST-NT, or GST-CT conjugated to glutathione-Sepharose beads. Proteins specifically bound to the beads (*B*) along with aliquots of the soluble brain extract (starting material (*SM*)) and equal amounts of the unbound material (void (*V*)) were subjected to SDS-PAGE and detected by Coomassie Blue staining. The migratory position of p50 (*asterisk*) as well as the mSos1 GST fusion proteins and GST alone are indicated on the *right*. The migratory positions of the molecular weight standards are indicated on the *left*.

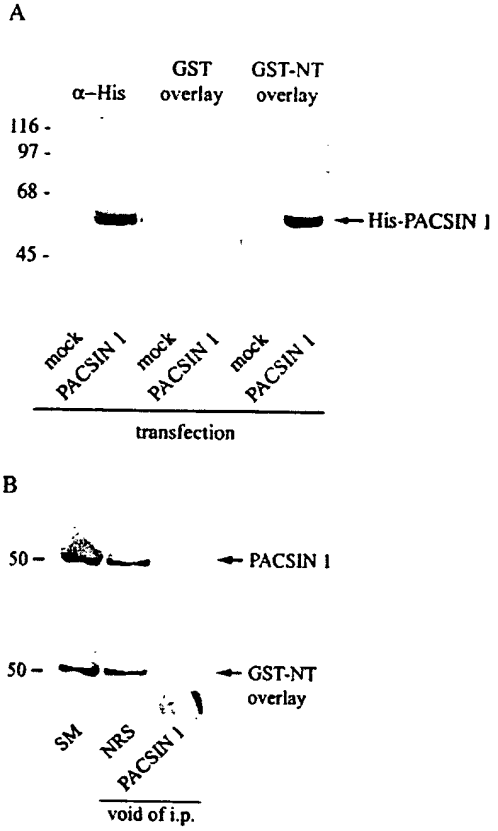


FIG. 4. **Confirmation of p50 as PACSIN 1.** *A*, cytosolic fractions of HEK-293 cells transfected with vector alone (*mock*) or with His_6-tagged PACSIN 1 (*PACSIN 1*) were resolved by SDS-PAGE and processed for Western blot with an anti-His_6 antibody (*α-His*) or were overlaid with GST or GST-NT. The migratory positions of His_6-PACSIN 1 and the molecular weight standards are shown on the *right* and *left*, respectively. *B*, an anti-PACSIN 1 serum (*PACSIN 1*) and normal rabbit serum (*NRS*) were coupled to protein A-Sepharose beads and incubated with a rat brain cytosolic extract. The unbound material (*void of i.p.*), along with an aliquot of the cytosolic extract (starting material (*SM*)) were processed for Western blot with the anti-PACSIN I antibody or were overlaid with GST-NT.

PACSIN 1 directly interacts with mSos1. The identity of p50 as PACSIN 1 was confirmed with the demonstration that immunodepletion of PACSIN 1 from brain extracts using an anti-PACSIN 1 antibody completely depletes p50, as determined by GST-NT overlay (Fig. 4*B*).

In addition to PACSIN 1, a second member of the PACSIN family, referred to as PACSIN 2, has been recently described (24, 25). In contrast to PACSIN 1, which is expressed exclusively in neurons, PACSIN 2 is expressed in brain and nonneuronal tissues (24, 25) (Fig. 5*A*). Because mSos1 is enriched in brain but is also expressed in nonneuronal tissues (9), we hypothesized that mSos-PACSIN 2 interactions may occur both within and outside the nervous system. To explore this question, we performed pull-down assays from Triton X-100-solubilized extracts prepared from brain, as well as from lung, which expresses high level of PACSIN 2 (Fig. 5*A*) (26). As expected, we detected binding of PACSIN 2 from both tissues to the mSos1 PRD (Fig. 5*B*). In fact, PACSIN 2 is likely to represent the 70-kDa band that was weakly detectable on the Coomassie blue stained pull-downs from brain extracts using mSos GST-NT (Fig. 3). Surprisingly, the level of PACSIN 2 recovered on the mSos1-PRD fusion protein was consistently greater when using brain *versus* lung extracts, even though PACSIN 2 was more abundant in the extracts from lung (Fig. 5, *A* and *B*). Moreover, a second, slightly smaller band that reacted with the PACSIN 2 antibody was detected in the GST-NT pull-downs from brain extracts but not from lung extracts (Fig. 5*B*). This protein may represent the short splice variant of PACSIN 2 previously described in rat tissues (25). Comparable amounts of the 70-kDa protein were pulled down from lung and brain extracts with an anti-PACSIN 2 antibody, suggesting that PACSIN 2 is equally accessible in both tissues (Fig. 5*B*). Thus, the differential binding of the long form of PACSIN 2 to the mSos1 PRD in brain compared with lung reveals a tissue-specific regulation of the interaction. The reason for this observation is currently unknown. However, it is possible that a tissue-specific posttranslational modification of PACSIN 2 alters its affinity for the mSos PRD.

To address whether the observed interactions are dependent on the classical SH3 domain binding interface, we generated point mutations in the SH3 domains of PACSIN 1 and 2 that converted proline to leucine (P434L for PACSIN 1; P478L for PACSIN 2). Comparable mutations in the *Caenorhabditis elegans* Grb2 homologue *sem-5* cause a lethal phenotype by preventing *sem*-5 interactions with its PRD-containing binding partners (35). Using wild-type and mutated PACSIN 1 and 2 expressed as GST fusion proteins, we performed pull-down assays from brain extracts (Fig. 5*C*). mSos1 was found to interact with both wild-type fusion proteins, whereas the proline to leucine mutations abolished mSos1 binding to both PACSINs, demonstrating that the interactions are specifically mediated through the PACSIN SH3 domains.

To explore the potential interaction between PACSIN 1 and mSos1 *in situ*, we performed co-immunoprecipitation experiments from rat brain extracts. Immunoprecipitation of PACSIN 1 led to co-immunoprecipitation of mSos1 (Fig. 6). The interaction was specific, as no mSos1 precipitated in the presence of normal rabbit serum, and the abundant brain protein tubulin was not detected in the anti-PACSIN 1 immunoprecipitates (Fig. 6). Only a limited percentage of the total mSos1 in the brain extract co-immunoprecipitated with PACSIN 1. This is not surprising given that PACSIN 1 interacts through its SH3 domain with multiple binding partners, including the abundant brain proteins dynamin 1 and synaptojanin 1 (23). In

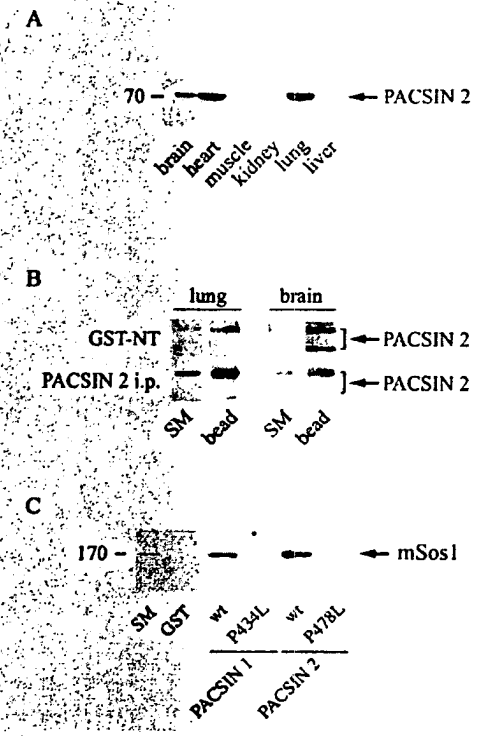


FIG. 5. **Specific interaction of mSos1 with the SH3 domains of PACSIN 1 and 2.** *A*, proteins of crude Triton-soluble extracts from various adult rat tissues (200 μg/tissue) were separated by SDS-PAGE, transferred to nitrocellulose, and processed for Western blot using antibodies against PACSIN 2. *B*, GST-NT conjugated to glutathione-Sepharose beads (*GST-NT*) and anti-PACSIN 2 antiserum precoupled to protein A-Sepharose beads (*PACSIN 2 i.p.*) were incubated with crude Triton-soluble extracts from rat lung and brain. Proteins specifically bound to the beads (*bead*) along with aliquots of the lung and brain extract (starting material (*SM*)) were subjected to SDS-PAGE and processed for Western blot using antibodies against PACSIN 2. *C*, a mouse brain extract was incubated with glutathione-Sepharose beads conjugated to GST alone or GST fused to full-length wild-type (*wt*) PACSINs or PACSINs with point mutations in their SH3 domains (PACSIN 1, proline 434 mutated to leucine (*P434L*); PACSIN 2, proline 478 mutated to leucine (*P478L*)). Material specifically bound to the beads was resolved by SDS-PAGE along with an aliquot of the brain extract (starting material (*SM*)) and processed for Western blot using antibodies against mSos1.

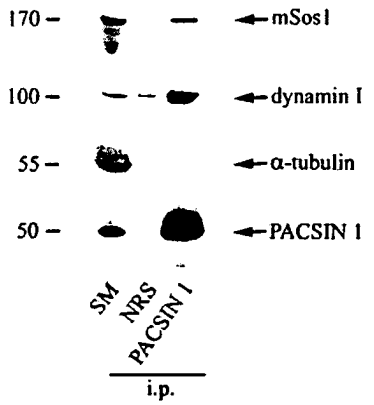


FIG. 6. **mSos1 interacts with PACSIN 1 *in vivo*.** Anti-PACSIN 1 antiserum (*PACSIN 1*) and normal rabbit serum (*NRS*), precoupled to protein A-Sepharose beads, were incubated with soluble extracts from rat brain. Material specifically bound to the beads (immunoprecipitation, *i.p.*) was resolved by SDS-PAGE along with an aliquot of the brain extract (starting material (*SM*)) and processed for Western blot using antibodies against mSos1, dynamin 1, tubulin, and PACSIN 1 as indicated.

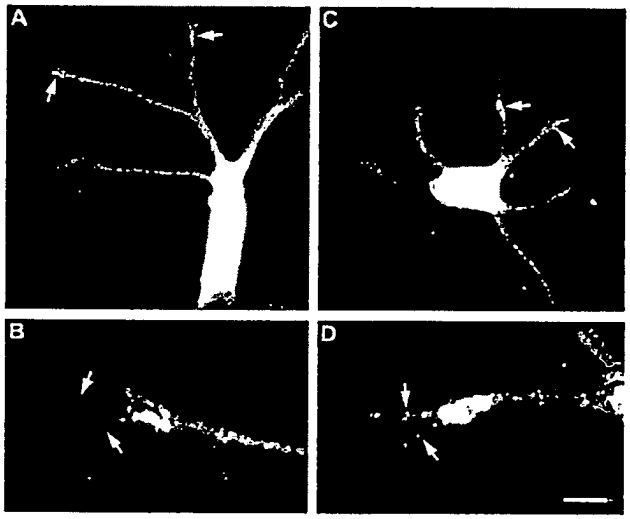


FIG. 7. **Co-expression of PACSIN 1 and mSos1 in growth cones and filopodia of cultured hippocampal neurons.** Confocal images of CA3 hippocampal neurons at 2 days *in vitro* stained for mSos1 (*A* and *B*) and PACSIN 1 (*C* and *D*). *Arrows* in *A* and *C* indicate growth cones. A higher magnification image of the growth cone reveals punctate staining along the neurites and filopodia (*arrows* in *B* and *D*). *Scale bar*, 10 μm in (*A* and *C*) and 1.5 μm in *B* and *D*.

fact, dynamin 1 was found to strongly co-immunoprecipitate with PACSIN 1 (Fig. 6). As the interactions between PACSIN 1 and its various SH3 domain-binding partners are likely to be competitive, the PACSIN-mSos interaction may be restricted to specific subcellular domains that are enriched for mSos1 relative to other PACSIN binding partners. Alternatively, the PACSINs may be at the core of large protein complexes in which they simultaneously interact with multiple binding partners. Consistent with the later possibility, it has been recently demonstrated that the PACSINs can self-associate to form homo- and hetero-oligomers (26).

To further examine the potential for interactions between PACSIN 1 and mSos1 *in situ*, we sought to determine whether the proteins were co-distributed in neurons. Immunofluorescence analysis of hippocampal neurons at 2 days *in vitro* with polyclonal antibodies against each protein revealed strong staining in the neuronal cell bodies, with fluorescent punctae observed along the length of the neurites and in growth cones (Fig. 7, *A* and *C*). Staining for both proteins extended into filopodia emanating from the growth cones (Fig. 7, *B* and *D*). To examine the localization within growth cones in more detail, we performed immunofluorescence analysis of primary rat dorsal root ganglia neurons maintained in culture for 1 day. Similar to hippocampal neurons, both PACSIN 1 and mSos1 were detected in dorsal root ganglia growth cones and were seen to extend into filopodia (Fig. 8). Interestingly, co-staining with phalloidin, which reveals filamentous actin, demonstrated that both mSos1 (Fig. 8*A*) and PACSIN 1 (Fig. 8*B*) were strongly co-localized with actin filaments at the plasma membrane and throughout the length of the filopodia. PACSIN has been demonstrated to localize at sites of high actin turnover, including filopodia and filopodial tips in nonneuronal cells (25). Further, overexpression of full-length PACSIN causes filopodia formation in an N-WASP-dependent manner, although the mechanism of this activation is unknown (25). The co-localization of mSos1 with PACSIN 1 in filopodia suggests that mSos1 may cooperate with PACSIN 1 in filopodia formation or function. This role may be particularly relevant during neuronal development, as actin-dependent filopodial dynamics are critical in the response of neuronal growth cones to extracellular guidance cues (36). Filopodia formation is dependent on N-WASP (37), which is activated by interactions with SH3 domains (38), as well as by binding to phosphatidylinositol

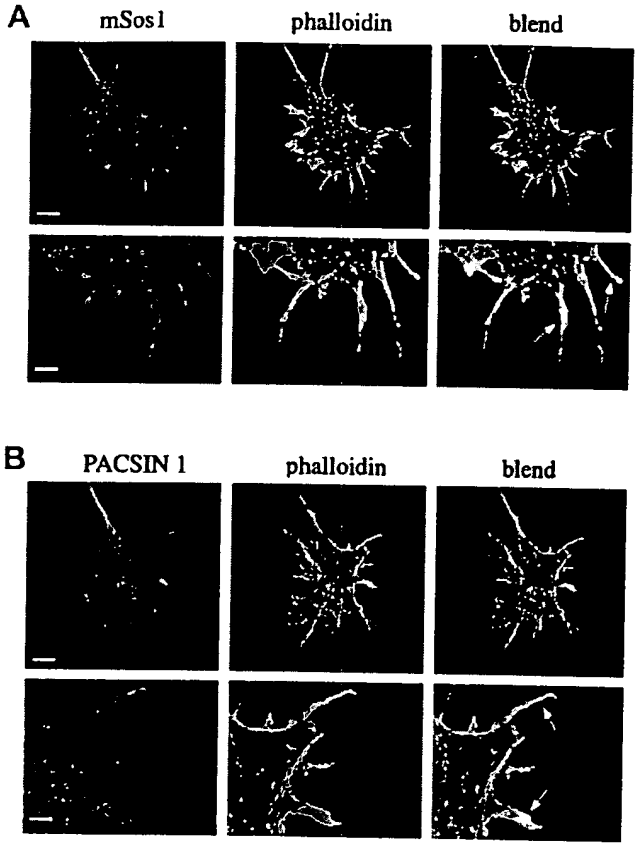


FIG. 8. **Co-localization of mSos1 and PACSIN 1 with actin in growth cones and filopodia of dorsal root ganglia neurons in culture.** Growth cones of E16 dorsal root ganglia neurons from an explant culture stained with an antibody against mSos1 (*A*) or PACSIN 1 (*B*) and co-stained with phalloidin-Alexa 488 (*phalloidin*) to demonstrate F-actin. *Arrows* point to filopodia demonstrating co-localization of mSos1 or PACSIN with actin. *Scale bar*, 5 μm in low power (*top three panels* of *A* and *B*) and 2 μm in high power (*bottom three panels* in *A* and *B*).

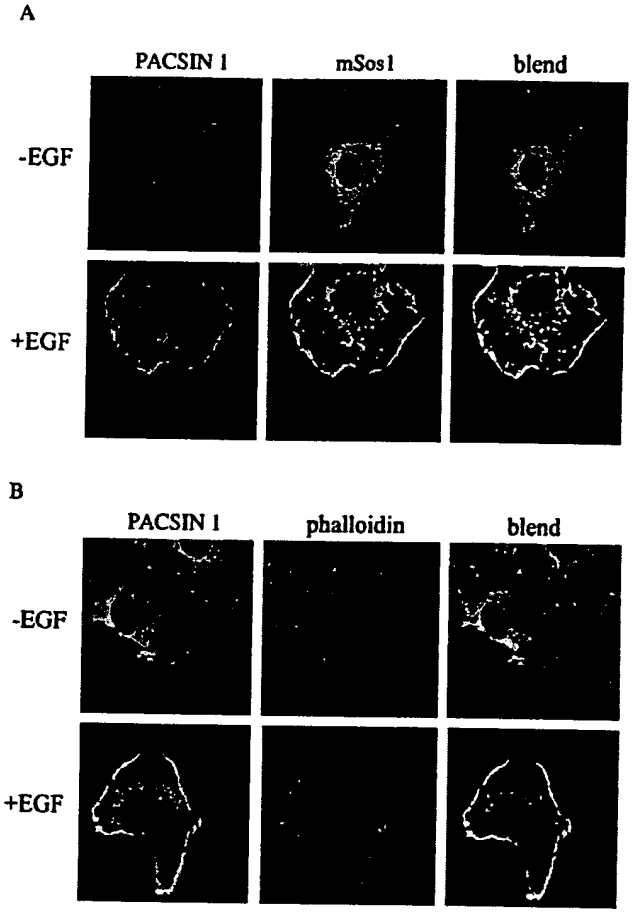


FIG. 9. **EGF treatment stimulates PACSIN 1/mSos1 co-localization in membrane ruffles.** COS-7 cells, co-transfected with His_6-tagged PACSIN 1 and FLAG-tagged mSos1 (*A*) or transfected with PACSIN 1 alone (*B*), were serum-starved and left untreated (–EGF) or were treated with 100 ng/ml EGF for 2 min (+*EGF*). PACSIN 1 was detected with an anti-PACSIN 1 rabbit antibody, and mSos1 was detected with anti-FLAG monoclonal antibody. F-actin was detected using phalloidin-tetramethylrhodamine isothiocyanate (*phalloidin*). Magnification, × 630.

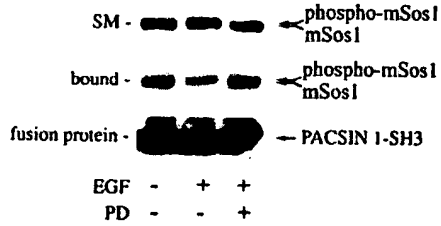


FIG. 10. **Regulation of the PACSIN 1-mSos1 interaction by phosphorylation.** Serum-starved COS-7 cells were treated with 100 ng/ml of EGF for 5 min at 37 °C (*EGF* +). When indicated, prior to the EGF treatment, cells were preincubated with PD-098059 (*PD* +). Lysates from cells were incubated with GST-PACSIN 1 SH3 coupled to glutathione-Sepharose beads, and material specifically bound to the beads was processed for SDS-PAGE, along with an aliquot of cell lysate (starting material (*SM*)). The migratory position of mSos1 and the upwardly shifted phosphorylated mSos1 (*phospho-mSos1*) and that of GST-PACSIN 1 SH3 domain are indicated on the *right*.

(4,5)bisphosphate and GTP-bound Cdc42 (39, 40). Thus, PACSIN may cause filopodia formation via SH3 domain-dependent stimulation of N-WASP. As activated Rac binds to phosphatidylinositol 4-phosphate 5-kinase, leading to phosphatidylinositol (4,5)bisphosphate production (41, 42), mSos could contribute to N-WASP stimulation via activation of Rac (17–19, 21).

We next investigated the possibility of direct co-localization of PACSIN 1 and mSos1 in actin-rich structures. Fibroblasts are well established to form membrane ruffles upon treatment with growth factors (43, 44). We therefore co-transfected COS-7 cells with FLAG-tagged mSos1 and His_6-tagged PACSIN 1 and examined their distribution before and after EGF treatment. Interestingly, both PACSIN 1 and mSos1 relocalized from a predominantly cytoplasmic distribution to become concentrated and co-localized at membrane ruffles following EGF treatment (Fig. 9*A*). Co-staining of PACSIN 1 transfected cells with anti-PACSIN 1 antibody and fluorescent phalloidin confirmed that the structures at which PACSIN 1 and mSos1 were co-localized were actin-rich membrane ruffles (Fig. 9*B*). Growth factor-induced ruffle formation is mediated by Ras-dependent Rac activation (43). Recent data suggest that mSos plays an important dual role in coupling Ras to Rac (18). Through its CDC25 homology domain, mSos activates Ras, which in turn activates phosphatidylinositol 3-phosphate kinase (45). The products of phosphatidylinositol 3-phosphate kinase catalytic activity stimulate the GEF activity of mSos toward Rac, causing Rac activation (17, 18). Indeed, mSos has been demonstrated to function directly in growth factor-induced membrane ruffle formation (18, 21). Our observation that full-length mSos1 targets to sites of Rac activation in response to EGF stimulation is consistent with its role in Rac-dependent actin reorganization. The co-localization of PACSIN 1 with mSos1 suggests that through direct protein interactions, PACSIN may regulate mSos function in ruffle

formation. In fact, PACSIN 1 interacts with mSos1 within the same region that mediates mSos1 interactions with Eps8 and E3b1 (18). The binding of these proteins to mSos stimulates mSos GEF activity toward Rac (18). Thus, PACSIN interactions with mSos may play a comparable modulatory role.

We next sought to determine whether the interaction between PACSIN 1 and mSos1 was subject to regulation in response to growth factor stimulation. Several laboratories have demonstrated that MEK-dependent feedback phosphorylation of mSos leads to Ras desensitization by inducing the dissociation of mSos from Grb2 and Shc (10, 28, 46–49). We thus examined mSos1-PACSIN 1 interactions following MEK-induced mSos1 phosphorylation. Treatment of COS-7 cells with EGF led to a small but highly reproducible upward shift in mSos1 mobility (Fig. 10). This shift is characteristic of MEK-induced mSos1 phosphorylation (28, 49) and was blocked by preincubation with the MEK inhibitor PD-098059 (Fig. 10). Interestingly, phosphorylated mSos1 showed less binding to the SH3 domain of PACSIN 1 than did nonphosphorylated mSos1 (Fig. 10; representative of three separate experiments). In contrast, the binding of N-WASP to the PACSIN 1 SH3 domain was identical under the various treatment conditions (data not shown). Thus, phosphorylation of mSos1 negatively regulates its interaction with PACSIN 1. Interactions between SH3 domain-containing proteins and their proline-rich binding partners are often regulated by phosphorylation. For example, phosphorylation of dynamin 1 and synaptojanin 1 reduces their interactions with the SH3 domains of the amphiphysins, allowing for regulation in their targeting to clathrin-coated pits (50).

The activation of Ras and the formation of membrane ruffles occurs rapidly following EGF treatment of COS-7 cells (both can be detected within 30 s).[2] In contrast, MEK-dependent feedback phosphorylation of mSos and the dissociation of the mSos-Grb2 complex is detectable only after several minutes (48). Thus, in parallel with the mSos-Grb2 interaction, it is likely that mSos and PACSIN are initially in association following their EGF induced translocation to membrane ruffles. Feedback phosphorylation via the Ras/MEK-dependent pathway, which appears to be a general mechanism to attenuate Ras signaling, would subsequently terminate the PACSIN-mSos interaction.

It is well established that Rac functions downstream of Ras in Ras-mediated signaling to alterations in the actin cytoskeleton (51). The ability of mSos to function as a GEF for Rac is central to the transduction of signals from Ras to Rac (16) and suggests the need for adaptor proteins involved in targeting mSos to sites of actin dynamics. Interestingly, Eps8, which activates the mSos GEF activity toward Rac following its indirect binding to mSos, also binds to actin (16). Thus, it is tempting to speculate that the PACSINs, which localize to sites of actin turnover and regulate actin cytoskeletal dynamics, may play a role in targeting mSos to sites of actin dynamics. Alternatively, interactions between PACSIN and mSos may allow for coordinated activities of the two proteins in regulation of the actin cytoskeleton. Future experiments will be aimed at testing these various hypotheses.

Acknowledgments—We thank Drs. Regis Kelly and Jeff Pessin for providing important reagents used in this study and acknowledge the contribution of the Howard Hughes Medical Institute Biopolymer Laboratory and the W. M. Keck Foundation Biotechnology Resource Laboratory at Yale University for mass spectrometry analysis.

REFERENCES

1. Campbell, S. L., Khosravi-Far, R., Rossman, K. L., Clark, G. J., and Der, C. J. (1998) *Oncogene* **17**, 1395–1413
2. Baltensperger, K., Kozma, L. M., Cherniack, A. D., Klarlund, J. K., Chawla, A., Banerjee, U., and Czech, M. P. (1993) *Science* **260**, 1950–1952
3. Buday, L., and Downward, J. (1993) *Cell* **73**, 611–620
4. Chardin, P., Camonis, J. H., Gale, N. W., van Aelst, L., Schlessinger, J., Wigler, M. H., and Bar-Sagi, D. (1993) *Science* **260**, 1338–1343
5. Egan, S. E., Giddings, B. W., Brooks, M. W., Buday, L., Sizeland, A. M., and Weinberg, R. A. (1993) *Nature* **363**, 45–51
6. Li, N., Batzer, A., Daly, R., Yajnik, V., Skolnik, E., Chardin, P., Bar-Sagi, D., Margolis, B., and Schlessinger, J. (1993) *Nature* **363**, 85–88
7. Rozakis-Adcock, M., Fernley, R., Wade, J., Pawson, T., and Bowtell, D. (1993) *Nature* **363**, 83–85
8. Leprince, C., Romero, F., Cussac, D., Vayssiere, B., Berger, R., Tavitian, A., and Camois, J. H. (1997) *J. Biol. Chem.* **272**, 15101–15105
9. Tong, X.-K., Hussain, N. K., de Heuvel, E., Kurakin, A., Abi-Jaoude, E., Quinn, C. C., Olson, M. F., Marais, R., Baranes, D., Kay, B. K., McPherson, P. S. (2000) *EMBO J.* **19**, 1263–1271
10. Tong, X.-K., Hussain, N. K., Adams, A., O'Bryan, J. P., and McPherson, P. S. (2000) *J. Biol. Chem.* **275**, 29894–29899
11. Di Guglielmo, G. M., Baass, P. C., Ou, W.-J., Posner, B. I., and Bergeron, J. J. M. (1994) *EMBO J.* **13**, 4269–4277
12. Haugh, J. M., Huang, A. C., Wiley, H. S., Wells, A., and Lauffenburger, D. A. (1999) *J. Biol. Chem.* **274**, 34350–34360
13. McPherson, P. S. Kay, B. K., and Hussain, N. K. (2001) *Traffic*, **2**, 375–384
14. Bowtell, D., Fu, P., Simon, M., and Senior, P. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 6511–6515
15. Whitehead, I. P., Campbell, S., Rossman, K. L., and Der, C. J. (1997) *Biochim. Biophys. Acta* **1332**, F1–F23
16. Scita, G., Tenca, P., Frittoli, E., Tocchetti, A., Innocenti, M., Giardina, G., and Di Fiore, P. P. (2000) *EMBO J.* **19**, 2393–2398
17. Das, B., Shu, X., Day, G. J., Han, J., Krishna, U. M., Falck, J. R., and Broek, D. (2000) *J. Biol. Chem.* **275**, 15074–15081
18. Scita, G., Nordstrom, J., Carbone, R., Tenca, P., Giardina, G., Gutkind, S., Bjarnegard, M., Betsholtz, C., and Di Fiore, P. P. (1999) *Nature* **401**, 290–293
19. Lanzetti, L., Rybin, V., Grazia Malabarba, M., Christoforidis, S., Scita, G., Zerial, M., and Di Fiore, P. P. (2000) *Nature* **408**, 374–377
20. Hall, A. (1998) *Science* **279**, 509–514
21. Nimnual, A. S., Yatsula, B. A., and Bar-Sagi, D. (1998) *Science* **279**, 560–563
22. Plomann, M., Lange, R., Vopper, G., Cremer, H., Heinlein, U. A. O., Scheff, S., Baldwin, S. A., Leitges, M., Cramer, M., Paulsson, M., and Barthels, D. (1998) *Eur. J. Biochem.* **15**, 201–211
23. Qualmann, B., Roos, J., DiGregorio, P. J., and Kelly, R. B. (1999) *Mol. Biol. Cell* **10**, 501–513
24. Ritter, B., Modregger, J., Paulsson, M., and Plomann, M. (1999) *FEBS Lett.* **454**, 356–362
25. Qualmann, B., and Kelly, R. B. (2000) *J. Cell Biol.* **148**, 1047–1062
26. Modregger, J., Ritter, B., Witter, B., Paulsson, M., and Plomann, M. (2000) *J. Cell Sci.* **113**, 4511–4521
27. Ramjaun, A. R., Micheva, K. D., Bouchelet, I., and McPherson, P. S. (1997) *J. Biol. Chem.* **272**, 16700–16706
28. Zhao, H., Okada, S., Pessin, J. E., and Koretzky, G. A. (1998) *J. Biol. Chem.* **273**, 12061–12067
29. McPherson, P. S., Takei, K., Schmid, S. L., and De Camilli, P. (1994) *J. Biol. Chem.* **269**, 30132–30139
30. McPherson, P. S., Czernik, A. J., Chilcote, T. J., Onofri, F., Benfenati, F., Greengard, P., Schlessinger, J., and De Camilli, P. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 6486–6490
31. Baranes, D., Lederfein, D., Huang, Y. Y., Chen, M., Bailey, C. H., and Kandel, E. R. (1998) *Neuron* **21**, 813–825
32. Lamarche-Vane, N., and Hall, A. (1998) *J. Biol. Chem.* **273**, 29172–29177
33. de Heuvel, E., Bell, A. W., Ramjaun, A. R., Wong, K., Sossin, W. S., and McPherson, P. S. (1997) *J. Biol. Chem.* **272**, 8710–8716
34. Roos, J., and Kelly, R. B. (1998) *J. Biol. Chem.* **273**, 19108–19119
35. Clark, S. G., Stern, M. J., and Horvitz, H. R. (1992) *Nature* **356**, 340–344
36. Zheng, J. Q., Wan, J. J., and Poo, M. M. (1996) *J. Neurosci.* **16**, 1140–1149
37. Miki, H., Sasaki, T., Takai, Y., and Takenawa, T. (1998) *Nature* **391**, 93–96
38. Carlier, M. F., Nioche, P., Broutin-L'Hermite, I., Boujemaa, R., Le Clainche, C., Egile, C., Garbay, C., Ducruix, A., Sansonetti, P., and Pantaloni, D. (2000) *J. Biol. Chem.* **275**, 21946–21952
39. Rohatgi, R., Ho, H. Y., and Kirschner, M. W. (2000) *J. Cell Biol.* **150**, 1299–1310
40. Prehoda, K. E., Scott, J. A., Mullins, R. D., and Lim, W. A. (2000) *Science* **290**, 801–806
41. Tolias, K. F., Hartwig, J. H., Ishihara, H., Shibasaki, Y., Cantley, L. C., and Carpenter, C. L. (2000) *Curr. Biol.* **10**, 153–156
42. Glogauer, M., Hartwig, J., and Stossel, T. (2000) *J. Cell Biol.* **150**, 785–796
43. Ridley, A. J., Paterson, H. F., Johnston, C. L., Diekmann, D., and Hall, A. (1992) *Cell* **70**, 401–410
44. Bretscher, M. S., and Aguado-Velasco, C. (1998) *Curr. Biol.* **8**, 721–724
45. Rodriguez-Viciana, P., Marte, B. M., Warne, P. H., and Downward, J. (1996) *Philos. Trans. R. Soc. Lond.-Biol. Sci.* **351**, 225–232
46. Cherniack, A. D., Klarlund, J. K., Conway, B. R., and Czech, M. P. (1995) *J. Biol. Chem.* **270**, 1485–1488
47. Rozakis-Adcock, M., Van der Geer, P., Mbamalu, G., and Pawson, T. (1995) *Oncogene* **11**, 1417–1426
48. Waters, S. B., Holt, K. H., Ross, S. E., Syu, L. J., Guan, K. L., Saltiel, A. R., Koretzky, G. A., and Pessin, J. E. (1995) *J. Biol. Chem.* **270**, 20883–20886
49. Corbalan-Garcia, S., Yang, S.-S., Degenhardt, K. R., and Bar-Sagi, D. (1996) *Mol. Cell. Biol.* **16**, 5674–5682
50. Slepnev, V. I., Ochoa, G. C., Butler, M. H., Grabs, D., and De Camilli, P. (1998) *Science* **281**, 821–824
51. Van Aelst, L., and D'Souza-Schorey, C. (1997) *Genes Dev.* **11**, 2295–2322

[2] S. Wasiak, C. C. Quinn, B. Ritter, E. de Heuvel, D. Baranes, M. Plomann, and P. S. McPherson, unpublished data.

Hum Genet (1997) 100: 114–122

ORIGINAL INVESTIGATION

Russell L. Margolis · Meena R. Abraham
Shawn B. Gatchell · Shi-Hua Li · Arif S. Kidwai
Theresa S. Breschel · O. Colin Stine · Colleen Callahan
Melvin G. McInnis · Christopher A. Ross

cDNAs with long CAG trinucleotide repeats from human brain

Received: 12 January 1997 / Accepted: 5 March 1997

Abstract Twelve diseases, most with neuropsychiatric features, arise from trinucleotide repeat expansion mutations. Expansion mutations may also cause a number of other disorders, including several additional forms of spinocerebellar ataxia, bipolar affective disorder, schizophrenia, and autism. To obtain candiate genes for these disorders, cDNA libraries from adult and fetal human brain were screened at high stringency for clones containing CAG repeats. Nineteen cDNAs were isolated and mapped to chromosomes 1, 2, 4, 6, 7, 8, 9, 12, 16, 19, 20, and X. The clones contain between 4 and 17 consecutive CAG, CTG, TCG, or GCA triplets. Clone H44 encodes 40 consecutive glutamines, more than any other entry in the nonredundant GenBank protein database and well within the range that causes neuronal degeneration in several of the glutamine expansion diseases. Eight cDNAs encode 15 or more consecutive glutamine residues, suggesting that the gene products may function as transcription factors, with a potential role in the regulation of neurodevelopment or neuroplasticity. In particular, the conceptual translation of clone CTG3a contains 18 consecutive glutamines and is 45% identical to the C-terminal 306 residues of the mouse numb gene product. These genes are therefore candidates for diseases featuring anticipation, neurodegeneration, or abnormalities of neurodevelopment.

R. L. Margolis · M. R. Abraham · S. B. Gatchell · S.-H. Li
A. S. Kidwai · T. S. Breschel · O. C. Stine · C. Callahan
M. G. McInnis · C. A. Ross
Laboratory of Molecular Neurobiology,
Johns Hopkins University School of Medicine,
600 N. Wolfe Street, Baltimore, MD 21287, USA

R. L. Margolis (✉) · C. A. Ross
Department of Psychiatry,
Johns Hopkins University School of Medicine, Meyer 2-181,
600 N. Wolfe Street, Baltimore, MD 21287, USA
Tel.: +1-410-614-4262; Fax: +1-410-955-8233;
e-mail: rmargoli@welchlink.welch.jhu.edu

Introduction

Trinucleotide repeat expansion mutation is now know to cause 12 diseases, most with neuropsychiatric features (Linblad and Schalling 1996; Paulson and Fischbeck 1996; Ross 1995; Zoghbi 1996). Seven of these are known as the type 1 disorders – spinocerebellar ataxia type 1 (SCA1, Orr et al. 1993), SCA2 (Imbert et al. 1996; Pulst et al. 1996; Sanpei et al. 1996), Machado-Joseph disease (MJD or SCA3, Kawaguchi et al. 1994), SCA6 (Zhuchenko et al. 1997), dentatorubral pallidoluysian atrophy (DRPLA, Koide et al. 1994; Nagafuchi et al. 1994), Huntington's disease (HD, Huntington's Disease Collaborative Research Group 1993), and spinal and bulbar muscular atrophy (SBMA, La Spada et al. 1991). Each is caused by a $(CAG)_n$ expansion in an open reading frame, resulting in an expanded glutamine repeat. The properties of the repeats in the other (type 2) expansion mutation diseases vary widely. Myotonic dystrophy is caused by a 3′ untranslated $(CTG)_n$ expansion (Brook et al. 1992; Fu et al. 1992; Mahadevan et al. 1992), the A and E forms of fragile X syndrome (Fu et al. 1991; Knight et al. 1993; Kremer et al. 1991; Verkerk et al. 1991) and some cases of Jacobsen's syndrome (Jones et al. 1995) result from 5′ untranslated region $(CCG)_n$ expansions, and Friedreich's ataxia is caused by an intronic $(GAA)_n$ expansion (Campuzano et al. 1996). Expandable trinucleotide repeats therefore are found in translated, transcribed but untranslated, and intronic regions; they may be G-C or A-T rich and range from minimal to highly variable in length in the normal population.

At least four lines of evidence indicate that additional disorders may arise from trinucleotide repeat expansion mutations. First, an antibody (IC2) that specifically recognizes expanded glutamine repeats detects an expansion segregating with SCA7 (Trottier et al. 1995). Second, indirect evidence of CAG expansion has been detected using rapid expansion detection (RED, Schalling et al. 1993) in a pedigree with SCA7, and less clearly in heterogeneous populations of patients with bipolar affective

disorder and schizophrenia (Linblad et al. 1996; Linblad and Schalling 1996; O'Donovan et al. 1995). Third, several neurodegenerative disorders, including SCA4, SCA5, SCA7, and familial Parkinson disease, are phenotypically similar to the type I expansion mutation disorders. Fourth, anticipation, the phenomenon of increasing phenotypic severity or decreasing age of onset in successive generations affected by a disease (McInnis 1996; Ross et al. 1993), is found in most of the expansion mutation diseases. Anticipation has been detected in a disparate group of other diseases, including affective disorder (Engstrom et al. 1995; McInnis et al. 1993; Nylander et al. 1994), schizophrenia (Chotai et al. 1995; Gorwood et al. 1996; Stober et al. 1995; Thibaut et al. 1995), autism (Stine 1993), familial Parkinsonism (Bonifati et al. 1995; Markopoulou et al. 1995; Payami et al. 1995; Plante-Bordeneuve et al. 1995), familial leukemias (Horwitz et al. 1996), Crohn's disease (Polito et al. 1996), Meniere's disease (Morrison 1995), torsion dystonia (LaBuda et al. 1993), rheumatoid arthritis (McDermott et al. 1996), facioscapulohumeral muscular dystrophy (Tawil et al. 1996), Holt-Oram syndrome (Newbury-Ecob et al. 1996), and familial spastic paraplegia (Raskind et al. 1997).

We have sought to identify candidate genes for these disorders by screening cDNA libraries for the presence of DNA fragments containing CAG, CCG, CCA, and AAT trinucleotide repeats (Li et al. 1993; Margolis et al. 1995 a, b). Our description of CTG-B37, a cDNA fragment with a highly polymorphic CAG repeat located within an open reading frame on chromosome 12, directly led to the finding that an expansion mutation within the CTG-B37 repeat causes DRPLA (Koide et al. 1994; Nagafuchi et al. 1994). This same strategy of screening cDNA libraries for trinucleotide repeats was later employed to identify the MJD gene (Kawaguchi et al. 1994) and the SCA6 gene (Zhuchenko et al. 1997). Screening genomic contigs for trinucleotide repeats was used to clone the gene for SCA2 (Pulst et al. 1996). Based on the repeats that expand to cause disease, repeats with the highest likelihood of undergoing expansion mutation consist of at least six consecutive CAG or CTG triplets in the transcribed portions of genes expressed in brain. To identify genes with these features, we have screened human adult frontal cortex and fetal brain cDNA libraries at high stringency for the presence of CAG or CTG repeats. We now report the identification and mapping of 19 of these cDNA fragments.

Materials and methods

cDNA cloning

Adult human frontal cortex (clones with CTG prefixes) and human fetal brain (all others) cDNA libraries (Strategene) was screened with either a $(CTG)_{20}$ or a $(CTG)_{15}$ radiolabeled oligonucleotide using standard techniques as previously described (Margolis et al. 1995a). In brief, approximately 200000–500000 plaques were screened in each of four sets of primary screens. Hybridization was at 52° (CTG and F clones) or 60° (H and L clones) with 50% formamide with a 70° (CTG and F clones) or 72° (H and L clones) wash at 0.2 × SSC/0.5% SDS for 15 min or until activity was below about 5 counts/s. After plaque purification, plasmids were excised using an in vivo excision procedure and purified.

Inserts were released by *Eco*RI digest, and Southern blots were prepared after electrophoresis through 1% agarose. Southern blots were hybridized as described for library screening, and positive inserts were sequenced using ABI automated technology until the repeat was detected or until a blastn search of the GenBank nonredundant, dbest, or dbsts libraries indicated that the clone had been previously described. Clones H32, 44, and 45 represent consensus double-stranded sequences of multiple clones. The contigs were formed in part by other clones identified in the library screen described above and in part by rescreening the fetal brain library with 45mer oligonucleotides antisense to sequence immediately adjacent to the repeat of each clone. H45: 5′ CCTGGGCACTCATTGGAGTCATGGTACTTATCATGGGTGTCTGCT 3′; H44: 5′ AGGTGGAATGGAGATGAGTCCCTGACGCTGAAGGCTGAGCAGATG 3′ and GGCCGCCCGTTCACCTTGATGAGACCATTGCCGCGTTTGCAGTGC; H32: 5′ AGGTGTGATGGTTTTTCCAGGGAGCTGAACTCCTGTGGCTTGGGA 3′.

Mapping

Clones were assigned to a chromosome or a specific locus using a variety of methods as indicated in Table 1. The NIGMS panel 2 monochromosomal human-rodent hybrid cell lines were screened using PCR primers that spanned the repeat as indicated in Table 2 (Dubois and Naylor 1993). Radiation hybrid mapping was performed by Research Genetics on either the GeneBridge 4 or Stanford G3 Radiation Hybrid panels using primers listed in Table 2. For L69 and F28, primers amplifying a region flanking the repeat were used to avoid crosshybridization with hamster genomic DNA (L69: 5′ CTCCGAGTATCGTGACAGGTG 3′, 5′ AAGCAGAGGTGAGGGAGAGTC 3′, 59° annealing temperature, 157-bp product; CAGF28: 5′ GCTGAAGGTGTAAAACCG 3′, 5′ GAAAGTGACTCCAGAGAC 3′, 52° annealing temperature, 86-bp product). CTG3a was mapped by analysis of repeat polymorphism in CEPH pedigrees 1332, 1347, 1362, and 884 (Weissenbach et al. 1992) with the LINKAGE program (Ott 1991). As indicated, other clones were assigned to a locus either through a GenBank match with a previously mapped ESTs (expressed sequence tags) or through use of Gene Map (Schuler et al. 1996).

Polymorphism analysis

Variability of repeat length was determined by size separation of radiolabeled PCR products on a 6% denaturing polyacrylamide gel, as previously described (Margolis et al. 1995a). Template consisted of 40–100 ng human genomic DNA from either a set of unrelated CEPH parents or a panel of unrelated individuals with movement disorders of unknown etioloy, and reaction conditions were as described in Table 2. Allele lengths of additional individuals with movement disorders were tested for size variation in L69 using automated detection (ABI) of fluorescence-labeled PCR product.

Results

Table 1 summarizes the initial characterization of the cDNA clones identified by screening adult human frontal cortex or human fetal brain cDNA libraries with a $(CAG)_{15}$ or $(CAG)_{20}$ probe. Those clones for which data on both polymorphism and chromosomal location are available are included, and clones corresponding to known genes were excluded except where noted, leaving

Table 1 Repeat sequence, mapping, and homologies

Clone	Genbank accession	Longest repeat	Conceptual translation of repeat region	Mapping method[a]	Chromosomal location	Homology
CTG3a	U80758	6	Q_{18}	E	19q13.13–19q13.2 D19S224–D19S228	mouse numb (U70764)
CTG4a	U80744	8	L_4PL_9	A	1	
CTG7a	U80745	8	$Q_5LQ_8HLIKLHHQNQ_6LQRIAQLQLQ_{16}$	A	20	
CTG20a	U80749	13	$PQ_{23}P_3$ (or $A_{13}TA_{11}$)	A	6	
F28	U80735	7	Q_8, 54% Q in 68 residue region	B	7q36.3 qter of D7S2781	mcag32 (U23862)
H1	U80738	13	$(AQ)_{11}ASQASQ_{15}$ or L_{15}	D	12p12 D12S328–D12S89	
H3	U80747	5′= 8 3′= 8	5′: Q_{14} 3′: Q_{18}	C	4	2.119 (5′ repeat only)
H16	U80737	12	Q_{28}	D	20q13.13 D20S891–D20S109	
H26	U80750	4	$Q_2PQ_8PQ_4PQ_3PQPQPQ_4PQ_3PQ$	A	16	mcag19 (U23860)
H32	U80743	15	Q_{30}	B	12q24.33 qter of D12S367	
H38	L13744	10	S_{42}	C	9q22	AF-9 (L13744)
H39	U80740	17	none	B	Xp11.4 AFMA212ZG1–DXS6745	
H44	U80741	7	$Q_{40}HPGKQAKEQ_{10}$	B	6q14–15 D6S1291	
H45	U80742	7	$Q_{25}YHIQ_6ILRQ_{25}HQ_6GP(XQ)_5$	C	1	KIAA0192 (D83783)
L69	U80752	13	none	B	2q37.3 D2S2704	
L85	U80755	10	S_{10} or A_{15} or Q_{11}	A	8	
L114	U80765	6	Q_{11}	B	12q13.12 WI-7107-DS12S1363	
L234	U80751	9	none	B	2q21.2 AFMA153WBI	
L237	U80757	15	none	B	16q24.3 D16S520-WI-12410	

[a]Mapping methods: A, NIGMS panel 2; B, radiation hybrid panel; C, identity to mapped cDNA; D, Gene Map; E, linkage

a total of 20 repeat regions from 19 indepdendent cDNA clones. The number of consecutive triplets ranges from 4 to 17 (mean = 9.8). Of the 20 repeats, 16 appear to be located in extended open reading frames. Of these 16, conceptual translation indicates that 12 (and possibly 14) encode polyglutamine. In each case the CAG repeat is interrupted frequently by CAA triplets also encoding glutamine. Eight of these polyglutamine regions consist of 15 or more consecutive glutamines, including Q_{28} (H16), Q_{30} (H32), and two adjacent stretches of Q_{25} (H45). The longest consecutive glutamine repeat, Q_{40}, is present in H44 (Fig. 1).

The cDNAs correspond to loci on 12 different chromosomes, as determined by the methods indicated in Table 1: one each from chromosomes 4, 7, 8, 9, 19 and X; two each from chromosomes 1, 2, 6, 16, and 20; and three from chromosomes 12. Blastn and blastp homology searches (Altschul et al. 1990) of GenBank revealed that the repeats in F28 and H26 are present in short cDNA fragments termed mcag32 and mcag19 (Yamagata et al. 1996). H45 is identical to the 3′ portion of a longer cDNA fragment (D83783) isolated from the KG-1 human cell line. D83783 is ubiquitously expressed, and contains both a DNA topoisomerase II motif and an N terminal transmembrane domain. A short murine amino acid sequence termed the mopa box protein (A26892) is also similar to H45. The first repeat in H3 is contained in clone 2.119 (Neri et al. 1996). H38 is a portion of the human gene AF-9, identified in relationship to a t(9:11) translocation in which AF-9 fuses with ALL-1 to induce acute lymphocytic leukemia (Prasad et al. 1993). AF-9 may function in the transcriptional complex (Cairns et al. 1996; Welch and Drubin 1994). CTG3a is similar to the mouse numb gene (Zhong et al. 1996), with 44.6% amino acid identity and 85.3% amino acid conservation (ALIGN, Pearson and Lipman 1988) over the length of CTG3a, which extends from m-numb amino acid 267 to 585 (Fig. 2). A comparison of CTG3a with both *Drosophila* and mouse numb by Block Maker (Henikoff et al. 1995) suggested five possible blocks, corresponding to CTG3a amino acid residues 26–52, 57–68, 85–104, 123–133, and 256–264. The Gibbs alignment (Lawrence et al. 1993) suggested three blocks, corresponding to CTG3a amino acid residues 29–60, 188–226, and 256–278. A search of these alignments against the Blocks database using LAMA (Pietrokovsi 1996) revealed a match only to one block

Table 2 Polymorphism analysis

Clone	Longest nucleotide repeat	PCR primers	Anneal temperature	Product length (in bp)	No. of chromosomes tested	Allele bp: frequency	Heterozygosity
CTG3a	6	GGGCACTGGGGCCACTGAGG CCTGGGCACAAGCGGACACC	64°	147	32	141:0.03 147:0.66 150:0.31	62%
CTG4a	8	CGTCCCGCTGTCTTCTGCTTC AGGCGAACCCAGTCGTTCTCC	62°	112	28	100:0.25 109:0.68 112:0.07	21%
CTG7a	8	TGCAGCTCCAACAACAGCAAC CTGCTGCATCGGTGGCTGCTG	62°	101	40	86:0.02 101:0.80 104:0.18	20%
CTG20a	13	CACCATGTCGCTGAAGCCCC CGCCGGGCTTGCGGACATTG	61° (7-deaza-dGTP)	125	38	125:0.97 128:0.03	15%
F28	7	GAAAGTGACTCCAGAGAC GTAAGGTTGCTGAGAGAT	49°	145	32	145:1.00	0%
H1	13	GGAGACTGGAAGTGTGGT CTCCCAGGCATCACAGCA	47° (DMSO)	82	30	82:0.83 85:0.17	20%
H3	5′= 8	GCACAGCAGCAACAAAGG GTCCTAAGGGAGACCAAT	47° (DMSO)	118	32	118:0.84 124:0.03 127:0.06 130:0.03 133:0.03	31%
	3′= 8	ACAAAAACCCCTTGATGC CCTGGAGCTGTGGAGGTG	53°	102	32	102:1.00	0%
H16	12	GGGTGGCTATGATGATGC TGAAGACCTGGGGTTGCT	54°	120	30	120:0.53 123:0.43 129:0.03	60%
H26	4	ATCCTAATAACGGCACTTCC TTCCTGGTGGCGAAATCGTG	54°	224	32	215:0.19 224:0.81	25%
H32	15	ACATTTCCAGCTTCTCAGGCA TCTGTTGCTGCTGCTGCTGTT	57°	113	32	110:0.53 113:0.41 119:0.03 128:0.03	38%
H38	10	TTAATTTGTGAGGCTTTGAAAAAC CTAATAGGAGTATTCATACCAGCA	54°	174	32	174:0.88 177:0.06 180:0.06	25%
H39	17	CTCTGATTGCTTAGTGGACAA TTCAGTACATTGCTGCTGCTG	62°	101	26	89:0.23 98:0.69 101:0.08	31%
H44	7	CAGCAAGAGCAGTTACAT GCTTGCTTTCCAGGATGC	52°	167 (bp 424–590)	32	167:1.00	0%
		TGCTGCTGGAAGACAAGC GAAAGCAAGCGAAAGAGC	52° (DMSO)	79 (bp 581–659)	32	79:1.00	0%
H45	7	TGCTGCTGCCGGATGTGGTA GCAGGCGTCCTGTCAACA	59°	131 (bp 2388–2518)	28	131:1.00	0%
L69	13	AGGTGAGGGAGAGTCCAT TATGGGCTCCTGTTCCTG	55°	104	88	83:0.02 92:0.09 104:0.20 107:0.02 110:0.45 113:0.01 116:0.03 122:0.15 126:0.01	75%
L85	10	GCTCTCGGGCATTGATGTTA AGCAGCAGCGGTAGCAGTGG	58°	136	26	133:0.78 136:0.15 139:0.08	46%
L114	6	GGGACCCAGAACTTGGTTTG CGAAGCCTCGGACCTGATTC	57°	195	28	195:1.00	0%

Table 2 (continued)

Clone	Longest nucleotide repeat	PCR primers	Anneal temperature	Product length (in bp)	No. of chromosomes tested	Allele bp: frequency	Heterozygosity
L234	9	ACAAGAGGAAAAGGGACATAGC GCAAAACCAAATACCAGGTCTC	51°	94	32	124:0.88 127:0.09 130:0.03	14%
L237	15	GGTTCCCTGCACAGAAACCATC AGATGCCACCGCATTCGG	59°	120	26	96:0.04 99:0.04 108:0.04 117:0.58 120:0.08 123:0.23	33%

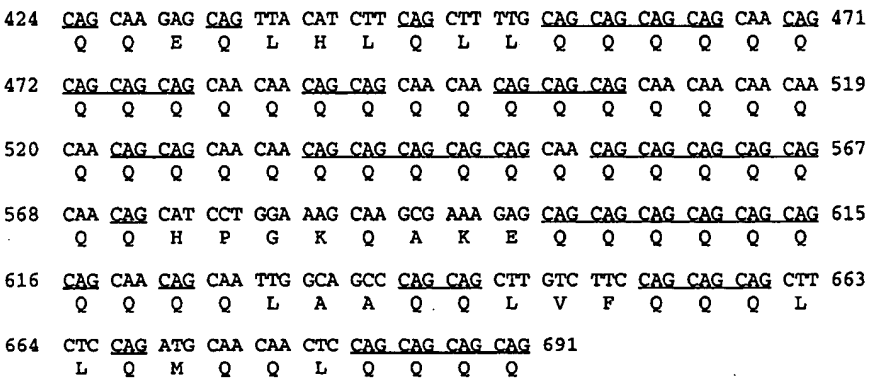


Fig. 1 Sequence and conceptual translation of the glutamine-rich region of H44. Forty consecutive glutamines are encoded by a CAG repeat interrupted in seven places by $(CAA)_{1-4}$

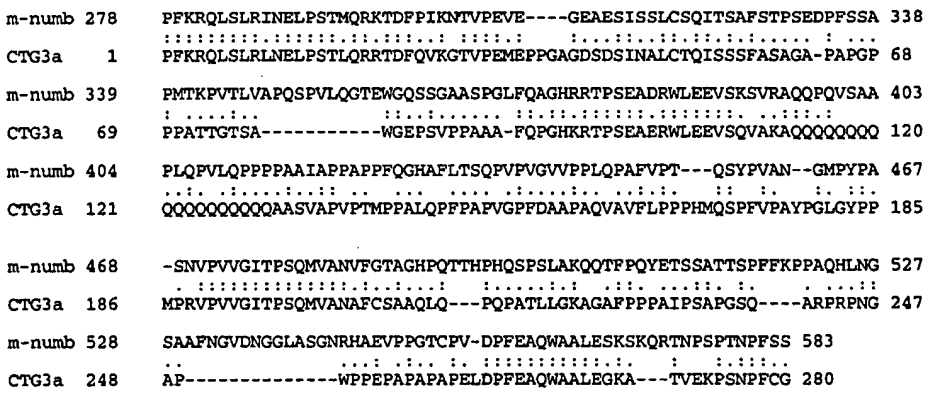


Fig. 2 Homology between the mouse numb gene product (*m-numb*) and the conceptual translation of *CTG3a*. There is 44.6% identity and 85.3% conservation between the sequences

formed from the synapsin family. This alignment is based on the abundance of glutamine and proline residues in the query block, and is of questionable biological significance. CTG3a and m-numb share a common PKC phosphorylation site (SLR at CTG3a amino acid residues 7–9) and three potential CamKinase II phosphorylation sites (CTG3a residues 28–31, 95–98, and 97–100). No other motif was detected and no other cDNA had significant homologies to other known genes at the DNA or protein level.

Table 2 summarizes the polymorphism analysis of this set of loci, as determined by radiolabeled PCR of CEPH individuals or individuals with movement disorders of unknown etiology. Heterozygosity ranged from 0% to 75%, with a mean of 25%. The 75% heterozygosity of L69 corresponded to the presence of 9 alleles, with the repeat presumably ranging in length from 6 to 20 triplets. Three other loci, CTG3a, H16, and L85 had fewer alleles but heterozygosity over 45%. Testing more individuals would probably have resulted in the detection of additional alleles. Several of the clones encoding long stretches of polyglutamine are not polymorphic; in each case the glutamine repeat is encoded by a combination of CAG and CAA codons.

To test the correlation of repeat length with the extent of heterozygosity, the modal repeat length was determined for each clone, and also for clones containing CAG repeats previously identified in our laboratory and for genes with CAG or CTG repeats in which expansion mutation is known to occur (Fig. 3). The repeat in ataxin-2, which expands to cause SCA2, is not highly polymorphic, and its ratio of heterozygosity to repeat length is similar to a number of the clones listed in Tables 1 and 2. All other CAG/CTG repeats associated with a disease are highly polymorphic with fairly long repeat lengths; the heterozygosity of L69, MAB21L1 (a gene previously identified in

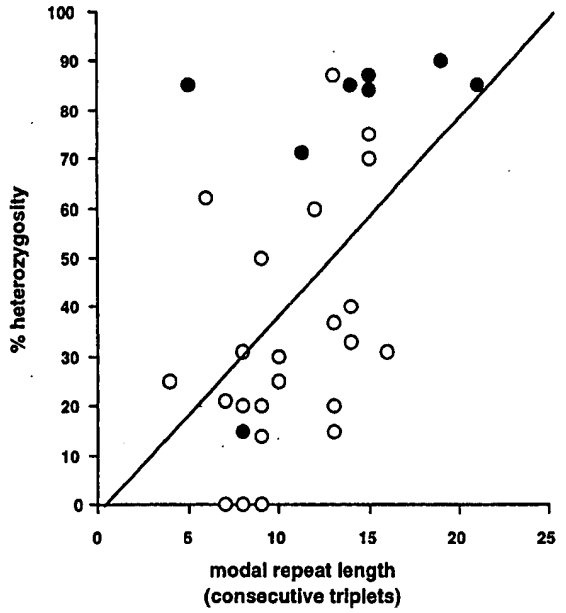


Fig. 3 The relationship between heterozygosity and modal repeat length. The modal repeat length and heterozygosity at the loci for CAG-containing clones identified in our laboratory (unfilled circles) were combined with similar information for the loci associated with CAG or CTG expansion mutations (filled circles). There is a marginal positive correlation between the two variables ($r^2 = 0.28$). Note that the repeat causing SCA2 is similar in length (8 triplets) and heterozygosity (15%) to many of the repeats listed in Tables 1 and 2

our laboratory and originally termed CAGR1, Margolis et al. 1996), and CTG-B33 (Li et al. 1993) are in a similar range.

Discussion

We now report the identification of 19 cDNA clones containing 20 CAG repeats. The repeats contain a mean of 10 consecutive triplets, with a maximum of 17. Most of the repeats are in open reading frames, contain the triplet CAA interspersed with longer runs of consecutive CAG triplets, and encode stretches of polyglutamine. The clones correspond to loci mapped to 12 different chromosomes and have a heterozygosity of up to 75%.

Four of the clones encode tracts of 25 to 40 consecutive glutamine residues. These are among the most glutamine-rich regions yet identified in the proteins of any species; in comparison, the ten most significant matches (blastp) in GenBank to a 40 glutamine query sequence consisted of proteins or protein fragments with 18 to 38 consecutive glutamines, only two of which were human. The longest stretches of consecutive glutamines were found in the human transcription factor TFIID (P20226, Q_{37} or Q_{38}, Imbert et al. 1994) and the yeast transcription factor SNF5 (P18480, Q_{37}), neither of which is homologous to any of the cDNA clones described here. The next longest glutamine repeats were present in the mouse mopa protein fragment (A26892, $Q_{33}YHIRQ_7MLRQ_{27}PHQ_5$) homologous to H45 and the uncharacterized yeast protein YM8520.13c (S54522, $Q_{30}HQ_5$). The remaining six proteins, with substantially fewer consecutive glutamines, were the *Drosophila* developmental regulatory protein notch (S09358, $Q_{13}HQ_{21}$), the fungal transcription factor white collar-1 (1480115, $HQHQ_{28}HQHQ_4$), a murine AT-motif binding factor (1585921, Q_{19}), a long human protein fragment nearly identical to H45 (D83783), the *Drosophila* ecdysone-induced protein E74A (B53225, Q_{18}), and a putative African malarial mosquito nucleic acid binding protein (S27770, Q_{20}). This list indicates that few glutamine repeats longer than 25 residues have yet been identified, and that proteins with long glutamine repeats, such as the gene products of H16, H32, H44, and H45, may function as transcription factors.

Clone H44, located on 6q14–15, is of particular interest since it encodes 40 consecutive glutamine residues and is present in the same length in all tested individuals. Repeat expansions resulting in as few as 21 consecutive glutamines in the α_{1A}-voltage-dependent calcium channel causes SCA6, and 35 consectuvie glutamines in ataxin-2 causes SCA2 (Imbert et al. 1996; Pulst et al. 1996; Sanpei et al. 1996). The minimal repeat length for disease onset is 35 or 36 glutamines in HD (Barron et al. 1993) and 40 glutamines in SBMA (La Spada et al. 1991) and SCA1 (Ranum et al. 1995). Normal alleles with more than 35 or 36 residues have not been detected in DRPLA or MJD, though the minimum expansion associated with disease is 49 glutamines for DRPLA and 68 glutamines for MJD (Kawaguchi et al. 1994; Koide et al. 1994; Nagafuchi et al. 1994). These findings, coupled with our data that a putative protein exists with 40 consecutive glutamines and the presence of at least 38 consecutive glutamines in TFIID, demonstrate that the development of an expansion mutation disease is not solely determined by the length of the glutamine repeat within a given gene. Other features of protein structure or expression must be relevant. The available data suggests, however, the presence of a threshold effect: it may be that a glutamine repeat of 60 or more residues in any protein leads to cell damage, whereas shorter stretches of glutamines are only toxic in a protein with certain as yet unidentified structural features or peculiarities of spatial or temporal expression. Conversely, shorter polyglutamine repeats may be inherently neurotoxic, with unknown aspects of structure, function, or expression serving a protective role in many, but not all, proteins. Comparative anylsis of TFIID and H44 with the α_{1A}-voltage-dependent calcium channel, ataxin-2, huntingtin, and the androgen receptor in cell toxicity assays (Ikeda et al. 1996) may help elucidate these features.

The CAG repeat-containing clones listed in Tables 1 and 2 can be considered candidate genes for neuropsychiatric disorders on the basis of both structure and function. From a structural standpoint, several of the repeats have either a high heterozygosity index (e.g., L69) or a fairly large number of alleles (H3 and L237), features of most of the repeats known to undergo expansion mutation. As depicted in Fig. 3, the ratio of modal repeat length to heterozygosity in many of the other cDNA clones is quite similar to ataxin-2; like ataxin-2, many of these cDNAs

have moderate stretches of CAG repeats interspersed with occasional CAA triplets. The presence of uninterrupted CAG repeats in expanded ataxin-1 and ataxin-2 demonstrates that CAA can mutate into CAG, leading to repeat length instability (Imbert et al. 1996; Orr et al. 1993; Pulst et al. 1996; Sanpei et al. 1996). A review of OMIM indicates that a number of diseases, though none with established anticipation, map near the loci listed in Table 1. Of particular interest, two developmental disorders, triphalangeal thumb-polysyndactyly syndrome (OMIM 190605) and sacral agenesis (OMIM 176450) map to 7q36 (clone F28), noninsulin-dependent diabetes mellitus 1 (OMIM 601283) and brachydactyly type E map to 2q37 (clone L69) progressive foveal dystrophy (OMIM 136550) maps to 6q14–16.2 (clone H44), and type 1 congenital nephrosis, Finnish type (OMIM 256300), maps to the same marker on 19q13 as CTG3a.

From the standpoint of function, we and others have postulated that developmental abnormalities may underlie at least some forms of schizophrenia, autism, and bipolar affective disorder (Keshavan et al. 1994; Margolis et al. 1994; Peterson 1995; Ross and Pearlson 1996; Weinberger 1995). Both *Drosophila* and murine numb appear to specify cell fate during early stages of neurodevelopment (Spana and Roe 1996; Zhong et al. 1996). The CTG3a gene product may share this capacity, though the extent of sequence identity (45%) between CTG3a and mouse numb indicates that the CTG3a gene product is not the human numb homolog but is instead a numb family member. Many of the other putative genes identified here may also have a specific role in the process of neuronal cell-fate determination, migration, or differentiation through regulation of transcription. A mutation, expansion or otherwise, would serve to alter this function. In particular, glutamine repeat length has been shown to modulate the efficieny of transcription (Chamberlain et al. 1994; Gerber et al. 1994). It is tempting to speculate that, though disease is induced by marked glutamine expansions, normal variations in glutamine repeat length may underlie normal variations in phenotype, modify the consequences of mutations in other genes, or influence the response to a nongenetic insult.

Acknowledgements The authors would like to thank Duane Bartley, Roxann Ashworth, and Dr. Alan Scott from the Johns Hopkins Genetics Core facility for assistance with sequencing, and John Kleiderlein, Kirsten Bottoms, and Hae Yoon for technical assistance. This work was supported in part by the National Alliance for Research on Schizophrenia and Depression, NIH MH02175-10A1, NIH NS34172, and NIH MH50763.

References

Altschul AF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Barron LH, Warner JP, Porteous M, Holloway S, Simpson S, Davidson R, Brock DJ (1993) A study of the Huntington's disease associated trinucleotide repeat in the Scottish population. J Med Genet 30:1003–1007

Bonifati V, Fabrizio E, Vanacore N, De Mari M, Meco G (1995) Familial Parkinson's disease: a clinical genetic analysis. Can J Neurol Sci 22:272–279

Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Davies J, Shelbourne P, Buxton J, Jones C, Juvonen V, Johnson K, Harper PS, Shaw DJ, Housman DE (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. Cell 68:799–808

Cairns BR, Henry NL, Kornberg RD (1996) TFG/TAF30/ANC1, a component of the yeast SWI/SNF complex that is similar to the leukemogenic proteins ENL and AF-9. Mol Cell Biol 16: 3308–3316

Campuzano V, Montermini L, Molto MD, Pianese L, Cossee M, Cavalcanti F, Monros E, Rodius F, Duclos F, Monticelli A, Zara F, Canizares J, Koutnikova H, Bidichandani SI, Gellera C, Brice A, Trouillas P, Demichele G, Filla A, Defrutos R, Palau F, Patel PI, Didonato S, Mandel JL, Cocozza S, Koenig M, Pandolfo M (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science 271:1423–1427

Chamberlain NL, Driver ED, Miesfelf RL (1994) The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. Nucleic Acids Res 22:3181–3186

Chotai J, Engstrom C, Ekholm B, son Berg ML, Adolfsson R, Nulander PO (1995) Anticipation in Swedish families with schizophrenia. Psychiatr Genet 5:181–186

Dubois BL, Naylor SL (1993) Characterization of NIGMS human/rodent somatic cell hybrid mapping panel 2 by PCR. Genomics 16:315–319

Engstrom C, Thornlund AS, Johansson EL, Langstrom M, Chotai J, Adolfsson R, Nylander PO (1995) Anticipation in unipolar affective disorder. J Affect Disord 35:31–40

Fu Y-H, Kuhl DPA, Pizzuti A, Pieretti M, Sutcliffe JS, Richards S, Verkerk AJMH, Holden JJA, Fenwick RG, Warren ST, Oostra BA, Nelson DL, Caskey CT (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. Cell 67:1047–1058

Fu Y-H, Pizzuti A, Fenwick RG, King J, Rajnarayan S, Dunne PW, Dubel J, Nasser GA, Ashizawa T, De Jong P, Wieringa B, Korneluk R, Perryman MB, Epstein HF, Caskey CT (1992) An unstable triplet repeat in a gene related to myotonic muscular dystrophy. Science 255:1256–1258

Gerber H-P, Seipel K, Georgiev O, Hofferer M, Hug M, Rusconi S, Schaffner W (1994) Transciptional activation modulated by homopolymeric glutamine and proline stretches. Science 263: 808–811

Gorwood P, Leboyer M, Falissard B, Jay M, Rouillon F, Feingold J (1996) Anticipation in schizophrenia: new light on a controversial problem. Am J Psychiatry 153:1173–1177

Henikoff S, Henikoff JG, Alford WJ, Pietrokovski S (1995) Automated contruction and graphical presentation of protein blocks from unaligned sequences. Gene 163:17–26

Horwitz M, Goode EL, Jarvik GP (1996) Anticipation in familial leukemia. Am J Hum Genet 59:990–998

Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72: 971–983

Ikeda H, Yamaguchi M, Sugai S, Aze Y, Narumiya S, Kakizuka A (1996) Expanded polyglutamine in the Machado-Joseph disease protein induces cell death in vitro and in vivo. Nat Genet 13:196–202

Imbert G, Trottier Y, Beckmann J, Mandel JL (1994) The gene for the TATA binding protein (TBP) that contains a highly polymorphic protein coding CAG repeat maps to 6q27. Genomics 21:667–668

Imbert G, Saudou F, Yvert G, Devys D, Trottier Y, Garnier J-M, Weber C, Mandel J-L, Cancel G, Abbas N, Durr A, Didierjean O, Stevanin G, Agid Y, Brice A (1996) Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. Nat Genet 14:285–291

Jones C, Penny L, Mattina T, Yu S, Baker E, Voullaire L, Langdon WY, Sutherland GR, Richards RI, Tunnacliffe A (1995) Association of a chromosome deletion syndrome with a fragile site with the proto-oncogene CBL2. Nature 376:145–149

Kawaguchi Y, Okamoto T, Taniwaki T, Aizawa M, Inoue M, Katayama S, Kawakami H, Nakamura S, Nishimura M, Akiguchi I, Kimura J, Narumiya S, Kakizuka A (1994) CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. Nat Genet 8:221–228

Keshavan MS, Anderson S, Pettegrew JW (1994) Is schizophrenia due to excessive synaptic pruning in the prefrontal cortex? The Feinberg hypothesis revisited. J Psychiatr Res 28:239–265

Knight SJL, Flannery AV, Hirst MC, Campbell L, Christodoulou Z, Phelps SR, Pointon J, Middletonprice HR, Barnicoat A, Pembrey ME, Holland J, Oostra BA, Bobrow M, Davies KE (1993) Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. Cell 74: 127–134

Koide R, Ikeuchi T, Onodera O, Tanaka H, Igarashi S, Endo K, Takahashi H, Kondo R, Ishikawa A, Hayashi T, Saito M, Tomoda A, Miike T, Naito H, Ikuta F, Tsuji S (1994) Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). Nat Genet 6:9–12

Kremer EJ, Pritchard M, Lynch M, Yu S, Holman K, Baker E, Warren ST, Schlessinger D, Sutherland GR, Richards RI (1991) Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. Science 252:1711–1714

LaBuda MC, Fletcher NA, Korczyn AD, Inzelberg R, Harding AE, Pauls DL (1993) Genomic imprinting and anticipation in idiopathic torsion dystonia. Neurology 43:2040–2043

La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature 352:77–79

Lawrence CE, Altschul SF, Boguski MS, Liu JS, Neuwald AF, Wooton JC (1993) Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. Science 262:208–214

Li S-H, McInnis MG, Margolis RL, Antonarakis SE, Ross CA (1993) Novel triplet repeat containing genes in human brain: cloning, expression, and length polymorphisms. Genomics 16: 572–579

Linblad K, Schalling M (1996) Clinical implications of unstalbe DNA repeat sequences. Acta Paediatrica 85:265–271

Linblad K, Savontaus M-L, Stevanin G, Holmberg M, Digre K, Zander C, David G, Benomar A, Nikoskelainen E, Trottier Y, Holmgren G, Ptacek LJ, Anttinen A, Brice A, Schalling M (1996) CAG repeat expansions in SCA7. Am J Hum Genet 59: A269

Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, Nelville C, Narang M, Barcelo J, O'Hoy K, Leblond S, Earle-MacDonald J, DeJong PJ, Wieringa B, Korneluk RG (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3′ untranslated region of the gene. Science 255:1253–1255

Margolis RL, Chuang D-M, Post RM (1994) Programmed cell death: implications for neuropsychiatric disorders. Biol Psychiatry 35:946–956

Margolis RL, Breschel TS, Li S-H, Kidwai AS, Antonarakis SE, McInnis MG, Ross CA (1995a) Identification and characterization of cDNA clones containing CCA trinucleotide repeats derived from human brain. Somat Cell Mol Genet 21:279–284

Margolis RL, Breschel TS, Li S-H, Kidwai AS, McInnis MG, Ross CA (1995b) Polymorphic (AAT)n trinucleotide repeats derived from a human brain cDNA library. Hum Genet 495–496

Margolis RL, Stine OC, McInnis MG, Ranen NG, Rubinsztein DC, Leggo J, Jones Brando LV, Kidwai AS, Loev SJ, Breschel TS, Callahan C, Simpson SG, Depaulo JR, McMahon FJ, Jain S, Paykel ES, Walsh C, DeLisi LE, Crow TJ, Torrey EF, Ashworth RG, Macke JP, Nathans J, Ross CA (1996) cDNA cloning of a human homologue of the *Caenorhabiditis elegans* cell fate-determining gene mab-21: expression, chromosomal localization and analysis of a highly polymorphic (CAG)n trinucleotide repeat. Hum Mol Genet 5:607–616

Markopoulou K, Wszolek ZK, Pfeiffe RF (1995) A Greek-American kindred with autosomal dominant, levodopa-responsive parkinsonism and anticipation. Ann Neurol 38:373–378

McDermott E, Khan MA, Deighton C (1996) Further evidence for genetic anticipation in familial rheumatoid arthritis. Ann Rheum Dis 55:475–477

McInnis MG (1996) Anticipation: an old idea in new genes. Am J Hum Genet 59:973–979

McInnis MG, McMahon FJ, Chase GA, Simpson SG, Ross CA, DePaulo JR (1993) Anticipation in bipolar affective disorder. Am J Hum Genet 53:385–390

Morrison AW (1995) Anticipation in Meniere's disease. J Laryngol Otol 109:499–502

Nagufuchi S, Yanagisawa H, Sato K, Shirayama T, Ohsaki E, Bundo M, Takeda T, Tadokoro K, Kondo I, Murayama N, Tanaka Y, Kikushima H, Umino K, Kurosawa H, Furukawa T, Nihei K, Inoue T, Sano A, Osamu K, Takahashi M, Yoshizawa T, Kanazawa I, Yamada M (1994) Expansion of an unstable CAG trinucleotide on chromosome 12p in dentatorubral and pallidoluysian atrophy. Nat Genet 6:14–17

Neri C, Albanese V, Lebre A-S, Holbert S, Saada C, Bougueleret L, Meier-Ewert S, Le Gall I, Millasseau P, Bui H, Giudicelli C, Massart C, Guillou S, Gervy P, Poullier E, Rigault P, Weissenbach J, Lennon G, Chumakov I, Dauseet J, Lehrach H, Cohen D, Cann HM (1996) Survey of CAG/CTG repeats in human cDNAs representing new genes: candidates for inherited neurological disorders. Hum Mol Genet 5:1001–1009

Newbury-Ecob RA, Leanage R, Raeburn JA, Young ID (1996) Holt-Oram syndrome: a clinical genetic study. J Med Genet 33: 300–307

Nylander PO, Engstrom C, Chotai J, Wahlstrom J, Adolfsson R (1994) Anticipation in Swedish families with bipolar affective disorder. J Med Genet 31:686–689

O'Donovan MC, Guy C, Craddock N, Murphy KC, Cardno AG, Jones LA, Owen MJ, McGuffin P (1995) Expanded CAG repeats in schizoprenia and bipolar disorder. Nat Genet 10:380–381

Orr HT, Chung M-Y, Banfi S, Kwiatkowski TJ, Servadio A, Beaudet AL, McCall AE, Duvick LA, Ranum LPW, Zoghbi HY (1993) Expansion of an unstable trinucleotide repeat in spinocerebellar ataxia type 1. Nat Genet 4:221–226

Ott J (1991) Analysis of human genetic linkage, 2nd edn. The Johns Hopkins University Press, Baltimore

Paulson HL, Fischbeck KH (1996) Trinucleotide repeats in neurogenetic disorders. Annu Rev Neurosci 19:79–107

Payami H, Bernard S, Larsen K, Kaye J, Nutt J (1995) Genetic anticipation in Parkinson's disease. Neurology 45:135–138

Pearson WR, Lipman DJ (1988) Improved tools for biological squence analysis. Proc Natl Acad Sci USA 85:2444–2448

Peterson BS (1995) Neuroimaging in child and adolescent neuropsychiatric disorders. J Acad Child Adolesc Psychiatry 34: 1560–1576

Pietrokovski (1996) Searching databases of conserved sequence regions by aligning protein multiple-alignments. Nucleic Acids Res24:3836–3845

Plante-Bordeneuve V, Taussig D, Thomas F, Ziegler M, Said G (1995) A clinical and genetic study of familial cases of Parkinson's disease. J Neurol Sci 133:164–172

Polito JM, Rees RC, Childs B, Mendeloff AI, Harris ML, Bayless TM (1996) Preliminary evidence for genetic anticipation in Crohn's disease. Lancet 347:798–800

Prasad R, Canaani O, Kamada N, Gale RP, Lange B, Crist WM, Nowell PC et al. (1993) Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute leukemia share sequence homology and/or common motifs. Proc Natl Acad Sci USA 90:4631–4635

Pulst S-M, Nechiporuk A, Nechiporuk T, Gispert S, Chen X-N, Lpes-Cendes I, Pearlman S, Starkman S, Orozco-Diaz G, Lunkes A, DeJong P, Rouleau GA, Auburger G, Korenberg JR, Figueroa C, Sahba S (1996) Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. Nat Genet 14:269–276

Ranum LP, Lundgren JK, Schut LJ, Ahrens MJ, Perlman S, Aita J, Bird TD, Gomez C, Orr HT (1995) Spinocerebellar ataxia type 1 and Machado-Joseph disease: incidence of CAG expansions among adult-onset ataxia patients from 311 families with dominant, recessive, or sporadic ataxia. Am J Hum Genet 57:603–608

Raskind WR, Pericak-Vance MA, Lennon F, Wolff J, Lipe HP, Bird TD (1997) Familial spastic paraparesis: evaluation of locus heterogeneity, anticipation, and haplotype mapping of the SPG4 locus on the short arm of chromosome 2. Am J Med Genet 74:26–36

Ross CA (1995) When more is less: pathogenesis of glutamine repeat neurodegenerative diseases. Neuron 15:493–496

Ross CA, Pearlson GD (1996) Schizophrenia, the heteromodal association neocortex and development: potential for a neurogenetic approach. Trends Neurosci 19:171–176

Ross CA, McInnis MG, Margolis RL, Li S-H (1993) Genes with triplet repeats: candidate mediators of neuropsychiatric disorders. Trends Neurosci 16:254–260

Sanpai K, Takano H, Igarashi S, Sato T, Oyake M, Sasaki H, Wakisaka A, Tashiro K, Ishida Y, Ikeuchi T, Koide R, Saito M, Sato A, Tanaka T, Hanyu S, Takiyama Y, Nishizawa M, Shimizu N, Nomura Y, Segawa M, Iwabuchi K, Eguchi I, Tanaka H, Takahashi H, Tsuji S (1996) Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. Nat Genet 14:277–284

Schalling M, Hudson TJ, Buetow KH, Housman DE (1993) Direct detection of novel expanded trinucleotide repeats in the human genome. Nat Genet 4:135–139

Schuler GD, et al (1996) A gene map of the human genome. Science 274:540–546

Spana EP, Roe CQ (1996) Numb antagonizes notch signaling to specify sibling neuron fates. Neuron 17:21–26

Stine OC (1993) Is autism associated with an X-linked trinucleotide repeat? Psychiatr Genet 3:129

Stober G, Franzek E, Lesch KP, Beckmann H (1995) Periodic catatonia: a schizophrenic subtype with major gene effect and anticipation. Eur Arch Psychiatry Clin Neurosci 245:135–141

Tawil R, Forrester J, Griggs RC, Mendell J, Kissel J, McDermott M, King W, Weiffenbach B, Figlewicz D (1996) Evidence for anticipation and association of deletion size with severity in facioscapulohumeral muscular dystrophy. Ann Neurol 39:744–748

Thibaut F, Martinez M, Petit M, Jay M, Campion D (1995) Further evidence for anticipation in schizophrenia. Psychiatry Res 59:25–33

Trottier Y, Lutz Y, Stevanin G, Imbert G, Devys D, Cancel G, Saudou F, Weber C, David G, Tora L, et al. (1995) Polyglutamine expansion as a pathological epitope in Huntington's disease and four dominant cerebellar ataxias. Nature 378:403–406

Verkerk AJMH, Pieretti M, Sutcliffe JS, Fu Y-H, Kuhl DPA, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang F, Eussen BE, Ommen van G-JB, Blonden LAJ, Riggins GJ, Chastain JL, Kunst CB, Galjaard H, Caskey CT, Nelson DL, Oostra BA, Warren ST (1991) Identification of a gene (FMR-1) containing CGG repeat coincident with a breakpoint cluster region exhibiting length variation in Fragile X syndrome. Cell 65:905–914

Weinberger DR (1995) From neuropathology to neurodevelopment. Lancet 346:552–557

Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, Lathrop M (1992) A second-generation linkage map of the human genome. Nature 359:794–801

Welch MD, Drubin DG (1994) A nuclear protein with sequence similarity to proteins implicated in human acute leukemias is important for cellular morphogenesis and actin cytoskeletal function in *Saccharomyces cerevisiae*. Mol Biol Cell 5:617–632

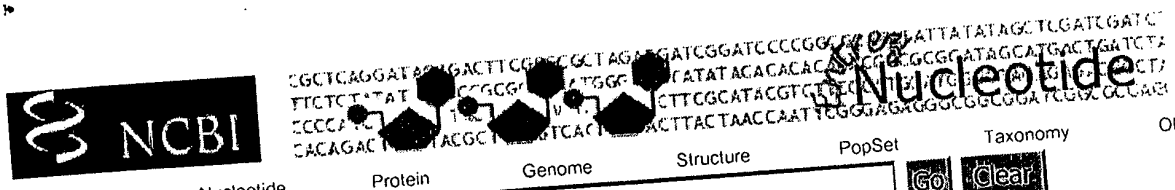
Yamagata K, Takeda J, Menzel S, Chen X, Eng S, Lim LR, Concannon P, Hanis CL, Spielman RS, Cox NJ, Bell GI (1996) Searching for NIDDM suscpetibility genes: studies of genes with triplet repeats expressed in skeletal muscle. Diabetologia 39:725–730

Zhong W, Feder JN, Jiang M-M, Jan LY, Jan YN (1996) Asymmetric localization of a mammalian numb homolog during mouse cortical neurogenesis. Neuron 17:43–53

Zhuchenko O, Bailey J, Bonnen P, Ashizawa T, Stockton DW, Amos C, Dobyns WB, Subramony SH, Zoghbi HY, Lee CC (1997) Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the a1A-voltage dependent calcium channel. Nat Genet 15:62–69

Zoghbi HY (1996) The expanding world of ataxins. Nat Genet 14:237–238

Exhibit A



Protein, Related Sequences, Taxonomy, LinkOut

1: AF087142 **Homo sapiens TED protein (TED) mRNA, complete cds**

```
LOCUS       AF087142     4042 bp    mRNA            PRI       01-OCT-1998
DEFINITION  Homo sapiens TED protein (TED) mRNA, complete cds.
ACCESSION   AF087142
VERSION     AF087142.1  GI:3676763
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4042)
  AUTHORS   Zonana,J. and Gault,J.
  TITLE     Isolation and characterization of a human X-linked gene mapping to
            the Xq13.1 region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4042)
  AUTHORS   Zonana,J. and Gault,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-1998) Molecular and Medical Genetics, Oregon
            Health Sciences University, L-103, 3181 S.W. Sam Jackson Road,
            Portland, OR 97201, USA
FEATURES             Location/Qualifiers
     source          1..4042
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xq13.1"
                     /clone="cosmid Ho4143"
     gene            1..4042
                     /gene="TED"
     CDS             48..1469
                     /gene="TED"
                     /codon_start=1
                     /product="TED protein"
                     /protein_id="AAC62086.1"
                     /db_xref="GI:3676764"
                     /translation="MFRGAWMWPGKDAAALTICCCCCCWAPRPSDKPCADSERAQRWR
                     LSLASLLFFTVLLADHLWLCAGARPRARELSSAMRPPWGAGRERQPVPPRAVLPVPPP
                     PPGEPSAPPGTCGPRYSNLTKAAPAAGSRPVCGGVPEPTGLDAACTKLQSLQRLFEPT
                     TPAPPLRPPDSPSRAPAEFPSAKKNLLKGHFRNFTLSFCDTYTVWDLLLGMDRPDSLD
                     CSLDTLMGDLLAVVASPGSGAWEACSNCIEAYQRLDRHAQEKYDEFDLVLHKYLQAEE
                     YSIRSCTKGCKAVYKAWLCSEYFSVTQQECQRWVPCKQYCLEVQTRCPFILPDNEEMV
                     YGGLPGFICTGLLDTSPKRLETKCCDVQWVSCEAKKKKKFKESEAPKTHQQQFHHSYF
                     HHYHQQYHHYHPHHDPPGRVSNKPALLPVSGGSRLSPSRIRLCVLVLMLLHTVVSFSS
                     NQGGGGLGLETLPALEEGLTREE"

BASE COUNT      885 a   1229 c   1129 g    799 t
ORIGIN
        1 tctggctcgc gcctgcctgt tccctccagc ccggaccccc ctgaaatatg ttcaggggcg
       61 cttggatgtg gcccgggaaa gacgccgccg cgctgactat ctgctgctgc tgctgctgct
      121 gggctcccag gccgagcgac aaaccttgcg ccgactccga gcgggcgcag cgatggcgac
```

```
 181 tgtccctggc gtccctgctc ttcttcaccg tgctgctcgc tgaccatctg tggctgtgcg
 241 cgggggcccg gccccgggcc agggagctga gcagcgccat gcggccccca tgggggccg
 301 gccgggagcg gcagccggtg cctcctcgcg cggtgctgcc cgtgccgccg ccgccgcccg
 361 gcgagcccag cgcgccccca ggcacctgcg gccccagata cagcaacctg accaaagccg
 421 cccccgccgc cggctctcgg ccggtctgcg gcggcgtccc agagcccacg gggctggacg
 481 cagcttgcac caaattgcaa tctttgcaga gacttttcga accgactact ccggcccccc
 541 ctctgcggcc ccctgactcc ccttcccgtg ccccggccga gttcccctcc gccaaaaaaa
 601 acttgctcaa aggccacttt cggaacttca ctctctcctt ttgcgacacc tacacggtct
 661 gggacttgct gctgggcatg gaccgccccg acagcctgga ctgtagcctg gacaccctga
 721 tgggggacct gctggccgtg gtggccagcc cgggctccgg ggcctgggag gcgtgtagca
 781 actgtatcga ggcgtaccag cggctggacc gacacgctca ggaaaaatat gacgagttcg
 841 acctcgtgct gcataaatac ttacaggcgg aagagtactc aatccggtcc tgcacgaaag
 901 gctgtaaggc tgtctacaag gcctggctgt gctcagaata cttcagcgtg acccagcagg
 961 aatgccagcg ctgggtgccc tgcaagcaat actgcctgga ggtgcagacc cggtgccccc
1021 ttatactccc cgacaatgag gaaatggtgt acggagggct ccctggcttt atctgtacag
1081 ggttgctgga tacttcacca aagcgtctgg aaaccaagtg ctgtgacgtg cagtgggtct
1141 cctgtgaggc gaagaagaag aagaagttca aggagtctga ggcccccaaa acccaccagc
1201 agcaattcca ccactcctat ttccaccact accaccaaca gtaccatcac taccaccccc
1261 atcatgatcc cccaggccgt gtcagcaaca agcccgccct gctgccggtc tctgggggct
1321 cccgcctcag ccctagcagg atccggctct gcgtccttgt tctcatgctc ctccataccg
1381 tggtgtcctt ctccagcaac cagggtggtg ggggattggg gctggagaca ctgcctgccc
1441 tagaggaggg cctgacacgg gaagagtgac agtagggagg gaggacagac ctccaccaca
1501 ctgacatcag ctccagctcc cccaggttgg gggggagggg ctcctcccat gggaggtgta
1561 ggataaggtg ggggcggggg aaatgggggа atgacacatc ccccccaacc taccccccacc
1621 ccaaaagcag ctccaacaga atggccagag ggcctcaggg agctgcaaaa ctcatccagg
1681 agaaaaaggc aggagcagca agtgacccct cccaagctcc catctatggg gcttagcaaa
1741 aaaagggagg aagtgggggc tggatatgcc cacccctgcc aaaagccctg accccaggga
1801 aggaggctgc tcacccagcc ttggccctgc agggaatgtt ggggggcaca gaggggagaa
1861 gctctttccc ccactccaca ttccttttgt tgccaagatc cttaattccc tcctgcccat
1921 atccctacag gcgacggcag acagtgcaat ggccctcctg ccacttcagc acaccttgcc
1981 ccacctggga ccatcacacg tgaagcacca ggctgggaga tgaggtgcac acagttgcag
2041 ctaggtcggg gccccagtta agctgtgccc caccacccta ggcaatgagg ggcaggaaag
2101 gggtacagaa tgaatggtga aagagagtgg agatacggaa gggggagagg agaggagaaa
2161 tgtggacaga gggcttgaag acatggccga ataggctact gccacccggc tttggggaga
2221 tgggcgataa gtgttgaggt aggctcgaga actgctcccc aaatcagtga gaattgcatt
2281 aggagcctct ggggaaagag cacaggaact gagttgctgg gcctcaaaac aaaaggctgc
2341 agcacatgtt agactgccgc tgtctgaaga ccatatctta cagaaggtgc acatactccc
2401 aagggccact cttgccctgg agatcttggc cttggggcaa agacatttcc atttccgatc
2461 tccatgggga ggagggactt aagcccaagt gggtcaggcc tggcccaggt ccttatactc
2521 ccctatcttg tcctagccca ggcctccatg aaggagaacc tggtggcatt gcccagagct
2581 ctccagagat gagcctcccc tatccccacc ctgtccccgt ccccagccta ccaaaggtat
2641 tcacaccttt ccatccctga ctctatgggt tactgtccca attgcgaaac acccatactc
2701 ctcttcaccc tgtagcaggg tcttctgctc cagtcacccc ggcctcttgc aagagaaagt
2761 ccagggactg gcaagaagac cctttgacta aagataatgt gctgtcatat cttggctgag
2821 aaactgttga gtcaggctag aggactggac aagagaagag gcattgagag ccccccttggg
2881 agccattagc cttgctctgg tctgctgtgt gggttgggga gtagggggga tcccacattc
2941 cctggacatc tgaatgctga tacgaatgga caggaaaggg agccagagct gtatacctgg
3001 gacaggccag ggagccttta gtgccagcag ggcctgtgtc ctgggatgtg atgctagtga
3061 gtgaatggga gggcgggcct ggagctgcgt gagagtgaaa ccaagaagcg gatggggaag
3121 ggagagctgg ccttggtcag cttagggatg aatggagaag gaatcattca tgggcctcgg
3181 ctaggtccca gactgattac atgagactgg ggagtccttc cctgccagtt ttctcattta
3241 ttacagccta tcccccaacc ctacccctgg gccctgccta cacactcccg tttggccctc
3301 agcactttgg gcctggtcac ataccccatg ccctagagtg ctcttttcag ctggggaggg
3361 gaagagtgct ggctcctttc tagacatgct tatgtataag taaaggggca ggtgcttggg
3421 ctccagagag accctaggta ctgccttcta actcctgcat cctcagaagg ggaaaaggga
3481 ctgggaagta aagggaaacc atatggcccc agggaatgga cccttggaga ctcccatcct
3541 ctcttccccc tcaccaagtg cccaggagac ccctggctca gaccagccca aggccttgcc
3601 cagtggcagg ggaaaggggc acctcactcc acctcaaagt catttgactg cccccatcca
3661 ccccaccacc ctaatcttcc accacctctc cggctgcccc tgcctcccat cacagacagc
3721 agagccgggc agctttctta tgccattttc tacactgtgc ttcatgagta ggactttctt
3781 gcactagttc ctatgactga gtctccaaac tggtttccta gtagtccccc atcccttcct
3841 cccttaccca gctatgattc agttgtctct gccctccctc ttaccctgcc tctgtgtttc
3901 ggtgagagtc tctgtatatg tactgctttc tgttttgttg cattgtgggg cagaggcctc
3961 tctgctcttg tttaacccca taaagaaata aatggtaaga cttgatgata aaaaaaaaaa
```

```
     4021 aaaaaaaaaa aaaaaaaaaa aa
//
```

Exhibit B

Am. J. Hum. Genet. 43:075–085, 1988

X-linked Hypohidrotic Ectodermal Dysplasia: Localization within the Region Xq11-21.1 by Linkage Analysis and Implications for Carrier Detection and Prenatal Diagnosis

Jonathan Zonana,*† Angus Clarke,* Mansoor Sarfarazi,* Nicholas S. T. Thomas,* Kim Roberts,* Kathi Marymee,† and Peter S. Harper*

*Institute of Medical Genetics, University of Wales College of Medicine, Cardiff; and †Department of Medical Genetics and Crippled Children's Division, Oregon Health Sciences University, Eugene

Summary

X-linked hypohidrotic ectodermal dysplasia (H.E.D.) is a disorder of abnormal morphogenesis of ectodermal structures and is of unknown pathogenesis. Neither relatively accurate carrier detection nor prenatal diagnosis has been available. Previous localization of the disorder by linkage analysis utilizing restriction-fragment polymorphisms, by our group and others, has placed the disorder in the general pericentromeric region. We have extended our previous study by analyzing 36 families by means of 10 DNA probes at nine marker loci and have localized the disorder to the region Xq11-Xq21.1, probably Xq12-Xq13. Three loci—DXS159 ($\hat{\theta}$ = .01, $\hat{z}$ = 14.84), PGK1 ($\hat{\theta}$ = .02, $\hat{z}$ = 13.44), and DXS72 ($\hat{\theta}$ = .02, $\hat{z}$ = 11.38)—show very close linkage to the disorder, while five other pericentromeric loci (DXS146, DXS14, DXYS1, DXYS2, and DXS3) display significant but looser linkage. Analysis of the linkage data yields no significant evidence for nonallelic heterogeneity for the X-linked form of the disorder. Both multipoint analysis and examination of multiply informative meioses with known phase establish that the locus for H.E.D. is flanked on one side by the proximal long arm loci DXYS1, DXYS2, and DXS3 and on the other side by the short arm loci DXS146 and DXS14. Multipoint mapping could not resolve the order of H.E.D. and the three tightly linked loci. This order can be inferred from published data on physical mapping of marker loci in the pericentromeric region, which have utilized somatic cell hybrid lines established from a female with severe manifestations of H.E.D., and an X/9 translocation (breakpoint Xq13.1). If one assumes that the breakpoint of the translocation is within the locus for H.E.D. and that there has not been a rearrangement in the hybrid line, then DXS159 would be proximal to the disorder and PGK1 and DXS72 would be distal to the disorder. Both accurate carrier detection and prenatal diagnosis are now feasible in a majority of families at risk for the disorder.

Introduction

X-linked hypohidrotic ectodermal dysplasia (H.E.D.) involves developmental defects in a number of ectodermal derivatives, including teeth, hair, and eccrine sweat glands, and can cause significant early morbidity and mortality (Freire-Maia and Pinheiro 1984; Clarke et al. 1987*a*). The exact frequency of the disorder is unknown, but a birth prevalence of 1/100,000 males and a mutation rate of $1–5 \times 10^{-6}$ have been estimated (Stevenson and Kerr 1967). Heterozygous females manifest overt signs of the disorder in only 70% of cases, and this has made accurate carrier detection in at-risk females problematic (Pinheiro et al. 1981). Analyses of sweat-pore numbers and distribution and radiographic measurement of tooth size in at-risk females can aid in carrier

Received November 18, 1987; revision received February 18, 1988.

Address for correspondence and reprints: Jonathan Zonana, M.D., Crippled Children's Division, University of Oregon, 901 East 18th Street, Eugene, OR 97403.

0002-9297/88/4301-0011$02.00

identification, but the study of sweat pores, despite various proposed methods, has proved quite variable and is subject to potential misclassification (Frias and Smith 1968; Nakata et al. 1980; Kleinebrecht et al. 1981; Happle and Frosch 1985). Prenatal diagnosis of the disorder has been reported once utilizing a skin biopsy obtained by fetoscopy, but the accuracy of this procedure has not been validated; nor is the procedure widely available (Arnold et al. 1984).

Recently, three groups of investigators confirmed a broad pericentromeric localization for the disorder, but without tight linkage to any single marker (Kolvraa et al. 1986; MacDermot et al. 1986; Clarke et al. 1987*b*). To further localize the disorder and examine the question of nonallelic genetic heterogeneity, the present study undertook to extend our previous work by the incorporation of 12 new families, nine American and three British, to give a total of 36 families. Seven new DNA probes were used from a total of nine marker loci analyzed.

Material and Methods

Family Studies

The 26 British families were ascertained by one of the authors (A.C.) writing to regional genetic centers and to medical and dental specialists. Consent to contact the families was obtained from the medical or dental practitioners involved. Samples and clinical information on one Swiss family were provided by Albert Schinzel.

Nine American families were ascertained through the cooperation of the National Foundation for Ectodermal Dysplasias. These families either had previously agreed to be contacted for research purposes or had read about the study in the foundation's newsletter. Informed consent for blood sampling and participation in the study was obtained in accordance with guidelines approved by the Human Research Review Board of the Oregon Health Sciences University.

Affected males and possible carrier females in all of the British families were personally examined by one of the authors (A.C.), and affected males had the typical findings of hypodontia, hypotrichosis, and hypohidrosis (Clarke et al. 1987*a*). Affected males in five of the nine American families were personally examined by one of us (J.Z.), and confirmation of the diagnosis in the others was obtained by both personal telephone interviews and, when necessary, review of clinical records. Potential carrier females in the American families were also personally interviewed by one of the authors as to manifestations of ectodermal dysplasia, but personal examination by one of us was not always possible.

Women were scored as heterozygous for H.E.D. either on genetic grounds (obligate carriers) or by the presence of abnormal dentition (manifesting carriers). Strict minimal criteria of dental abnormalities were selected for classification as a carrier and required the absence of two or more deciduous or permanent teeth (excluding third molars), absence of two caniniform deciduous or permanent incisors, or marked microdontia of at least four teeth of either dentition (Nakata et al. 1980). In a previous study utilizing the same minimal criteria for dental abnormalities, 78% of the British obligate carrier females had abnormal dentition (Clarke et al. 1987*a*). Systematic measurement of tooth size on dental radiographs was not performed. Many of the women who fit the dental criteria also had additional physical manifestations of H.E.D., such as thin scalp hair or patchy body hair, but these findings were not utilized since they were difficult to quantify.

Although examination of the pattern of sweating on the whole back for evidence of Lyonization had been performed previously on a majority of the possible British carriers, neither these data nor palmar sweat-pore counts were utilized in classification (Clarke et al. 1987*a*). Clarke et al. (1987*a*) had found that 97% of both the obligate and manifesting carriers (as classified by dental criteria) displayed abnormal sweat patterns. However, these tests were excluded to minimize the possibility of misclassification due to subjective interpretation. At-risk females without abnormal dentition were classified as possible carriers and were not used in the linkage analysis.

Southern Blot Analysis

High-molecular-weight DNA was extracted, by means of published procedures, from blood previously stored at −20 C or −70 C (Kunkel et al. 1977). DNA samples (3–5 μg) were digested to completion for 6–24 h by using 2–4 units of the relevant restriction endonucleases/μg DNA and were then separated by electrophoresis on 0.7%–0.9% agarose gels at 1.3–1.6 V/cm for 18–24 h. Transfer was then performed to nylon filters, either Hy-Bond® N (Amersham) or GeneScreen Plus® (duPont), according to standard manufacturer's procedures.

Hybridization of the filters with $^{P}32$-labeled probes was performed at 65 C for 24 h in 6 × SSC, 5 ×

Denhardt's solution, 0.5%–1% SDS, 7%–10% dextran sulfate, and 20 μg herring-sperm DNA/ml. DNA probes were labeled by either nick-translation, using a commercial kit (Amersham), or by random oligomer extension to give specific activities of 10^8–10^9 cpm/μg DNA (Feinberg and Vogelstein 1983). Posthybridization washes were at 65 C, twice at 2 × SSC and 0.1% SDS for 10–20 min and subsequently at increasing stringencies (1%–0.1% SSC with 0.1% SDS), depending on the probe. Autoradiography onto Fuji X-ray film was performed at −70 C with an intensifying screen for 1–7 days.

Some initial difficulties were encountered with the use of the phosphoglycerate kinase 1 (PGK1) cDNA probe pHPGK-7e, which detects either a 6-kb or both a 2-kb and a 4-kb fragment on *Pst*I digestion (Hutz et al. 1984). PGK1 is part of a multigene family, and an autosomal band overlays the 6-kb X-derived fragment, which necessitates a double digest with *Xba*I and *Pst*I to selectively remove the autosomal fragment. Despite this treatment a doublet of similarly sized fragments persisted at around 6 kb in a few individuals, but the slightly larger autosomal fragment could be either removed or clearly distinguished from the X-linked band following a longer period (24 h) of enzymatic digestion and electrophoresis for 24 h on a 0.7% agarose gel. The typing of all apparent heterozygous females was repeated using this second method to ensure that no homozygous females had been mistyped. Subsequent to these studies we obtained the genomic pSPT/PGK probe containing the 5′-flanking region of the PGK1 locus, which is also polymorphic (*Bgl*I) but does not have any overlying autosomal fragments (Keith et al. 1986; Vogelstein et al. 1987). Unfortunately the genomic *Bgl*I polymorphism was in complete linkage disequilibrium with the cDNA *Pst*I polymorphism (96 individuals tested), but it was simpler to use than the cDNA probe.

Linkage Analysis

The nine pericentromeric RFLP loci used in this study included three loci located on Xp (DXS7, DXS146, and DXS14) and six loci located on Xq (DXS159, PGK1, DXS72, DXYS1, DXYS2, and DXS3). The DXS7 locus was not tested on the nine American families, since no significant linkage was apparent after analysis of the British families. Ten DNA probes from these loci were utilized and are listed in table 1, which also includes information on probe localization, restriction enzymes utilized, and allele size and frequencies. Allele frequencies were derived from our 36 families and were calculated from 93–112 unrelated X chromosomes. Two-point linkage analyses to determine the maximum-likelihood recombination distances (θ_{max}) and the lod scores at θ_{max} ($\hat{z}$) between the main locus (HED) and the DNA probe loci were carried out using MLINK from the LINKAGE package (version 3.5) (Lathrop et al. 1984). The confidence intervals (CI) were calculated as all values of θ for which the lod score was within 1 unit of the maximum (Conneally et al. 1985). Two-point analyses were initially carried out utilizing data on all heterozygous females, including manifesting females, and on normal and affected males. The analyses were then repeated with a subset of data utilizing only the offspring of the 49 heterozygous females who were obligate carriers, both possible and manifesting carriers being excluded, to test whether there had been any systematic misclassification of females as heterozygotes on the basis of physical examination alone.

Multipoint analyses were performed for HED and the marker loci by using ILINK (Lathrop et al. 1984). The analyses utilized the full set of data, including all heterozygous females. We utilized a fixed order for marker loci whenever such an order was clearly established in the literature (e.g., DXS7-DXS14-DXYS1). We then performed a series of four-point analyses utilizing marker loci alone to determine, when possible, the position of the other marker loci. Subsequent to this, a series of four-point analyses were performed utilizing the HED locus and different combinations of three marker loci. In all the analyses the relative likelihoods (odds) and probabilities of the alternative orders were calculated.

The question of possible heterogeneity was addressed by utilization of two programs, HOMOG and HOMOG2 (1987 version), developed by J. Ott to test for homogeneity (admixture) (Ott 1983, 1985). The first program, HOMOG (test I), carries out a homogeneity test under the following alternative hypothesis: two family types, one with linkage between HED and the marker locus and the other without linkage. The second program, HOMOG2 (test II), is an extension of the homogeneity test, with the following alternative hypothesis: two family types, both with linkage and one with theta 1 and the other with theta 2, where theta 1 < theta 2 < .5. In both of these programs the tests of one hypothesis against another are carried out as likelihood-ratio tests of heterogeneity (H2) versus homogeneity (H1), with the *P* values calculated from the asymptotic χ^2 distribution

Table 1

RFLP Loci

Locus (Probe), Location, and Enzyme	Alleles (kb)	Frequencies	Source (Reference)
DXS7 (L1.28) Xp11.3:			
TaqI	12.0	.74	Pearson
	9.0	.26	(Wieacker et al. 1984)
DXS146 (pTAK8a) Xp11-q12:			
XbaI	5.0	.40	Kruse
	3.3	.60	(Kolvraa et al. 1986)
DXS14 (p58.1) Xp11-cen:			
MspI	4.0	.75	Latt
	2.5	.25	(Bruns et al. 1984)
DXS159 (cpX73) Xq11-q12:			
PstI	5.5	.59	Pearson
	1.6	.41	(Arveiler et al. 1987*a*)
PGK1 (pHPGK-7e) Xq13:			
PstI/XbaI	6.0	.21	Orkin
PstI	4.0/2.0	.79	(Hutz et al. 1984)
PGK1 (pSPT/PGK) Xq13:			
BglI	12.0	.30[a]	Riggs
	5.0	.70	(Vogelstein et al. 1987)
DXS72 (pX65H7) Xq21.1:			
HindIII	7.2	.54	Schmeckpeper
	.7	.46	(Schmeckpeper et al. 1985)
DXYS1 (pDP34) Xq21.31:			
TaqI	12.0	.35	Page
	11.0	.65	(Page et al. 1982)
DXYS2 (7b) Xq13-Xq22:			
PstI	12.0	.29	Weissenbach
	9.0	.71	(Koenig et al. 1985)
DXS3 (p19.2) Xq21.3-q22:			
TaqI	5.1	.38	Bruns
	2.95/2.15	.62	(Aldridge et al. 1984)

[a] Estimated on the basis of published data, owing to insufficient study data.

of the likelihood-ratio statistic. One should note that the tests of homogeneity are one sided and that the P values are one-half those associated with χ^2. These analyses were carried out between HED and all nine marker loci. The input to the programs consisted of lod scores calculated for the values of theta ranging from .0 to .49 (50 points).

Results

The results of the two-point linkage analyses for the full set of data, including all of the heterozygous females, are displayed in table 2. In table 3, analyses of the subset of data utilizing obligate carrier females only (method B) are compared with the first set of analyses (method A). Comparison of the results from the full set of data and from the subset of data shows some variation in the θ_{max} and z_{max} between HED and each loci, but the θ_{max} values of each are close to one another and well within the CI of the other analysis. This result would generally validate our classification methods in both the British and American families.

Three marker loci localized to the Xq11-21.1 region demonstrated very close linkage with the HED locus (table 2). The DXS159 locus (Xq11-q12) had only one phase-unknown recombinant, producing a $\hat{z}$ of 14.84 at a $\hat{\theta}$ of .01. The single recombinant individual was an edentulous 60-year-old woman whose classification as a heterozygous carrier is uncertain, since it is based on a history of two missing permanent teeth (excluding the third molars), and her dental records were not available for review. Physical examinations of her and of her three offspring (two male, 1 female) were otherwise normal. Since in vari-

Table 2

Two-Point Linkage of HED and Marker Loci

Locus	No. of Recombinant Meioses/Total No. of Meioses for Phase		θ (in cM) at a Recombination Fraction of									θ_{max}	z_{max}
	Known	Unknown	.01	.02	.03	.04	.05	.10	.20	.30	.40		
DXS7	7/21	6/25	−14.59	−10.56	−8.28	−6.71	−5.54	−2.25	.09	.63	.46	.31	.64
DXS146 ...	5/16	4/41	−2.12	.33	1.64	2.50	3.10	4.42	4.24	2.90	1.30	.14	4.59
DXS14	1/13	2/34	8.10	8.73	8.98	9.08	9.09	8.55	6.43	3.83	1.28	.05	9.09
DXS159 ...	0/35	1/33	14.84	14.84	14.70	14.52	14.30	12.99	9.93	6.58	3.19	.01	14.84
PGK1	0/23	1/33	13.43	13.44	13.33	13.16	12.96	11.76	8.94	5.90	2.85	.02	13.44
DXS72	0/17	1/28	11.34	11.38	11.28	11.13	10.96	9.87	7.36	4.74	2.22	.02	11.38
DXYS1	1/18	2/31	6.95	7.60	7.87	7.99	8.02	7.62	5.86	3.72	1.64	.05	8.02
DXYS2	2/24	6/36	1.20	3.31	4.43	5.13	5.60	6.48	5.69	3.76	1.55	.11	6.50
DXS3	8/34	6/45	−9.42	−4.95	−2.49	−.85	.34	3.34	4.41	3.35	1.50	.18	4.44

ous surveys 1%–7% of the normal population have been found to be missing permanent teeth other than the third molars, the possibility exists that this woman is not heterozygous for H.E.D. (Bailit 1975; Nakata et al. 1980). This individual was also the only recombinant between the disease and the PGK1 locus. One cannot infer gene order from this meiosis, since it is a phase-unknown situation, and her carrier status is not certain.

Only one phase-unknown recombinant (one of the two sons of an obligate-carrier female) was present between HED and the DXS72 locus (Xq21.1), yielding a $\hat{\theta}$ of .02 with a $\hat{z}$ of 11.38. This meiosis was also recombinant between HED and the DXYS2 locus. The PGK1 locus, localized to Xq13, was assessed by the use of both a cDNA probe, pHPGK-7e, and a genomic probe (pSPT/PGK) from the 5′-flanking region of the gene. A $\hat{\theta}$ of .02 with a $\hat{z}$ of 13.44 was found, with only one phase-unknown recombinant (discussed above).

On two-point analysis, five other pericentromeric marker loci (DXS146, DXS14, DXYS1, DXYS2, and DXS3) showed significant but looser linkage to HED, with $\hat{\theta}$'s of 5–18 cM and $\hat{z}$'s of 4.44–9.09. Only the DXS7 locus showed no significant linkage to HED. Examination of multiply informative meioses (phase known) suggested that the DXYS1, DXYS2, and DXS3 loci are all distal to HED, while the short arm

Table 3

Two-Point Analysis of HED versus Marker Loci

Locus	Method A[a]			Method B[b]			
	θ_{max}	z_{max}	CI[c]	θ_{max}	z_{max}	CI[c]	No. of Informative Meioses
DXS7	.31	.639	...	.21	.96	...	26
DXS146	.14	4.593	.06–.26	.11	3.127	.03–.27	33
DXS14	.05	9.087	<.01–.13	.00	8.068	.00–.06	32
DXS159	.01	14.844	<.01–.07	.00	9.648	.00–.05	39
PGK1	.02	13.442	<.01–.07	.00	10.156	.00–.05	39
DXS72	.02	11.375	<.01–.08	.04	4.163	<.01–.18	22
DXYS1	.05	8.023	.01–.14	.07	3.391	<.01–.22	23
DXYS2	.11	6.504	.05–.21	.13	2.233	.04–.31	24
DXS3	.18	4.439	.10–.29	.15	3.455	.06–.29	43

[a] Utilizing all heterozygous females.
[b] Utilizing only obligate-carrier heterozygous females.
[c] = −1 Lod score.

Table 4

Multipoint Analyses

Order of Loci	Estimated $\hat{\theta}$ between Loci (cM)	−Ln(L)	Relative (L)	Probability (%)
A:				
DXS7-DXS146-DXS14-PGK1	17.5-11.5-3.5	378.41	55.15	92.86
DXS146-DXS7-DXS14-PGK1	13.3-17.5-3.8	381.25	3.24	5.45
DXS7-DXS14-DXS146-PGK1	17.8-5.4-10.0	382.42	1.00	1.68
B:				
DXS146-DXS14-HED-DXYS1	9.7-4.4-5.2	818.57	15.20	93.82
DXS146-HED-DXS14-DXYS1	13.8-3.4-4.6	821.29	1.00	6.17
C:				
DXS14-PGK1-HED-DXYS1	2.5-1.3-5.5	793.22	2,936.58	95.06
DXS14-HED-PGK1-DXYS1	4.6-1.4-4.3	796.35	128.38	4.16
HED-DXS14-PGK1-DXYS1	2.9-1.7-4.0	798.05	23.22	.75
DXS14-PGK1-DXYS1-HED	3.1-2.1-3.2	801.20	1.00	.03
D:				
DXS14-PGK1-HED-DXYS2	2.5-1.3-11.0	793.72	511447.39	94.51
DXS14-HED-PGK1-DXYS2	4.5-1.5-10.3	796.85	22471.43	4.15
HED-DXS14-PGK1-DXYS2	3.1-1.4-9.1	797.98	7259.02	1.34
DXS14-PGK1-DXYS2-HED	1.8-3.6-7.6	806.87	1.00	.00
E:				
DXS14-PGK1-HED-DXS3	2.4-1.2-16.9	831.89	94584105.0	78.27
DXS14-HED-PGK1-DXS3	4.4-1.4-15.0	833.52	18531823.0	15.34
HED-DXS14-PGK1-DXS3	3.5-1.5-13.8	834.40	7725213.3	6.39
DXS14-PGK1-DXS3-HED	1.7-5.7-13.8	850.26	1.0	.00
F:				
DXS14-DXS72-HED-DXYS1	3.5-1.6-5.0	807.28	58.56	51.16
DXS14-HED-DXS72-DXYS1	4.3-1.5-4.2	807.38	53.20	46.47
DXS72-DXS14-HED-DXYS1	2.0-3.6-5.0	810.81	1.71	1.50
DXS14-HED-DXYS1-DXS72	4.4-3.6-2.1	811.35	1.00	.87
G:				
DXS146-DXS159-HED-DXYS1	12.8-1.6-5.0	827.82	4138.13	57.85
DXS146-HED-DXS159-DXYS1	13.8-1.1-4.7	828.14	3003.40	41.99
DXS159-DXS146-HED-DXYS1	1.5-7.7-5.0	833.76	10.89	.15
DXS146-HED-DXYS1-DXS159	13.8-3.4-3.7	836.15	1.00	.01

loci, DXS14 and DXS146, flanked the disorder on the other side.

Multipoint analyses, comparing the relative likelihoods (odds) and probabilities of the alternative orders of the loci, confirmed that HED is proximal to three Xq marker loci (DXYS1, DXYS2, and DXS3) and is bracketed by DXS146 and DXS14, which are located on proximal Xp (table 4). We first established that the DXS146 locus was distal to DXS14 (table 4A). Subsequent analysis demonstrated that both HED and PGK1 were between the DXS14 locus and the three long arm markers, DXYS1, DXYS2, and DXS3 (of the 12 orders calculated, only the four with the greatest relative likelihoods are shown; table 4B–4E). It was impossible to definitely resolve the order of HED and the three tightly linked loci PGK1, DXS159, and DXS72 (table 4C–4G). There were no recombinants between these three marker loci, and there were only single recombinants (phase unknown) between HED and each of these loci. There is a suggestion in the analysis that HED is distal to PGK1, with the relative odds ratios between the alternative orders in the range of 5–23:1, but the only multiply informative recombinant in the analysis is from the female whose carrier status is uncertain. The analysis revealed that the DXS72 locus is between DXS14 and DXYS1 and that DXS159 is between DXS146 and DXYS1 (table 4F, 4G). There were no recombinants between the DXS14 and DXS159 loci. Estimates derived from multipoint analyses of the re-

Table 5

Coupling of HED and Marker Loci for Three Tightly Linked Loci

Locus and Allele (kb)	Observed	Expected[a]	χ^2	*P*
DXS159:				
5.5	25	20	3.04	.08
1.6	9	14		
PGK1:				
6.0	9	7	.47	.49
4.0/2.0	26	28		
DXS72:				
7.2	19	19	.00	1.0
0.7	16	16		

[a] Derived from study sample.

combination distances between HED and the linked marker loci were within 1 cM of those calculated by two-point analyses.

Since three of the markers demonstrated tight linkage with HED, we examined the 36 families for evidence of linkage disequilibrium (table 5). No clear evidence for disequilibrium was present for any of the markers, which is consistent with the impression of a moderately high rate of new mutation among families with the disorder.

Results of the tests for heterogeneity between HED and the nine marker loci showed little evidence for significant heterogeneity with either of the two methods utilized for eight of the nine marker loci (table 6). In the first method (test I), which assumes two types of families, one with linkage and one without, five of the nine markers had a χ^2 value of 0.000, with the maximum possible probability of homogeneity (.5). The second method (test II), which assumes two types of families, both with linkage but with different values of theta, also yields generally similar results. The only exception to these results is seen in the data for the DXS3 locus. This exception may be due to the multiple comparisons performed in the analyses, possible heterogeneity of the position of the marker locus itself, or incorrect typing of this marker locus in a few families. The families that had a low posterior probability of linkage between HED and the DXS3 locus were informative for other marker loci and had a high posterior probability of linkage between HED and these loci.

Discussion

The first clue to the regional localization of H.E.D. on the X chromosome was the description of a female with severe manifestations of H.E.D., a case reported verbally at the first Human Gene Mapping conference by the late Dr. Peter Cook but not documented in the literature (Gerald and Brown 1974). The patient had previously been karyotyped for failure to thrive at 6 mo of age by other investigators, without recognition of her ectodermal dysplasia, and had a balanced X/9 translocation, with the breakpoint placed at Xq12 (Cohen et al. 1972; Shows and Brown 1974). We have had the opportunity to review the past records of this patient and have confirmed

Table 6

Tests for Heterogeneity between HED and Nine Marker Loci

Marker Locus (No. of Informative Families)	Test I		Test II	
	χ^2 (df = 1)	*P*	χ^2 (df = 2)	*P*
DXS7 (19)	.061	.4026	.119	.4712
DXS146 (26)	.000	.5000	.002	.4995
DXS14 (22)	.457	.2495	.457	.3978
DXS159 (25)	.000	.5000	.104	.4747
PGK1 (22)	.000	.5000	.003	.4796
DXS72 (27)	.042	.4183	.136	.4671
DXYS1 (23)	.000	.5000	.000	.5000
DXYS2 (20)	.000	.5000	.000	.5000
DXS3 (23)	4.661	.0154	6.347	.0209

Note.—Tests of heterogeneity assume that there are two types of families, as follows: for test I, one type is linked and the other type is unlinked (theta = .5); for test II, both types have linkage, one type having theta 1 and the other type having theta 2, where theta 1 < theta 2 < .50.

that she had typical manifestations of a male with H.E.D. (K. D. MacDermot, personal communication); we have also recently obtained a fibroblast cell line of this patient from the National Institute of General Medical Sciences' Human Genetic Mutant Cell Repository (GM0705) and, using high-resolution techniques, found the breakpoint actually to be at the junction of the Xq12 and Xq13.1 bands, with a karyotype of 46,X,t(X;9) (Xpter→Xq13.1::9p24→9pter; 9qter→9p24::Xq13.1→Xqter) (Zonana et al., in press). Her cell line, also referred to as the AnLy line, has been used extensively by a number of investigators in the construction of somatic cell hybrids for gene mapping purposes, apparently without realization of the patient's full disease phenotype (Shows and Brown 1974, 1975; Grzeschik and Siniscalco 1976; Wieacker et al. 1984; Oberle et al. 1986; Riddell et al. 1986; Arveiler et al. 1987*b*).

Another clue to the localization of the disorder is the existence of a similar X-linked disorder, the Tabby mutant, in the mouse (Blecher 1986). This mutant locus has been mapped close to the PGK-1 locus (Searle et al. 1987). If the Tabby locus is genetically homologous to the human X-linked HED locus, one would predict, on the basis of man-mouse X-chromosome homology, that the gene might be located on the proximal long arm near the PGK1 locus (Ohno 1973).

In 1986 MacDermot et al. suggested linkage of HED to the DXYS1 locus, with a $\hat{\theta}$ of .06, and a $\hat{z}$ of 2.7. Our present and past studies confirm linkage to the DXYS1 locus at a distance of 5 cM and demonstrate by multipoint analysis that this locus, as well as DXYS2 and DXS3, are all distal to HED (table 4). This analysis is consistent with reports of two males with both choroideremia, but without H.E.D., and interstitial deletions within the region Xq13-q22, in which one male was deleted for both the DXYS1 and DXS3 loci and the other for DXYS1, DXYS23, DXYS5, DXYS12, DXYS2, and DXS3 loci (Hodgson et al. 1987; Schwartz et al., in press). Kolvraa et al. (1986) previously suggested close linkage of the disorder to the DXS146 locus (Xp11-Xq12), with a $\hat{\theta}$ of .0, and a $\hat{z}$ of 2.41, but our present data indicate linkage at a distance of 14 cM, with the locus located distal to DXS14. We had previously demonstrated linkage of the HED locus to the DXS14 and DXS3 loci and suggested an order of DXS14-HED-DXS3, and this has been confirmed by our expanded data set, with only minor modifications of genetic distances (Clarke et al. 1987*b*).

The tightest linkage to HED was demonstrated with three marker loci (DXS159, PGK1, and DXS72) from the subregion Xq11-Xq21.1. None of these loci were deleted in any of our affected males, nor was clear linkage disequilibrium apparent (table 5). The lack of phase-known recombinants between HED and the DXS159 and PGK1 loci does not allow us to order the loci by linkage analysis. The third locus closely linked to HED is DXS72. The data suggest that this locus is distal to HED and proximal to DXYS1, on the basis of a doubly informative meiosis recombinant between HED and DXS72 and DXYS2 and on the basis of another informative meiosis recombinant between DXYS1 and both HED and DXS72. Although these are phase-unknown meioses, the suggested order agrees with information from a patient reported with both choroideremia without ectodermal dysplasia and an Xq21.1 interstitial deletion, deleted for both DXS72 and DXYS1 (Lesko et al. 1987). DXS72 has recently been assigned, by deletion mapping, to Xq21.1, proximal to DXYS1, which was placed at Xq21.3 (Yang et al. 1987).

The following most likely order of loci can be proposed on the basis of our linkage data: DXS7, DXS146 - DXS14 - (DXS159 / HED / PGK1 / DXS72) (DXYS1/DXYS2/DXS3). A major difficulty in determining the exact gene order for HED and the closely linked marker loci is the relative lack of genetic recombination in the pericentromeric region of the X chromosome. Although one may have close genetic linkage, there may still be a considerable physical distance between loci. Physical mapping of marker and disease loci will be necessary but has been hindered by the lack of any H.E.D. patients identified with either cytogenetic or probe deletions. However, one can utilize information from published somatic cell hybrid studies, which utilized the X-autosome translocation from the manifesting female with H.E.D. (the AnLy fibroblast line) to attempt to order the closely linked marker loci and HED.

Shows and Brown (1975) first showed, in a mouse hybrid, that the PGK1 locus mapped distal to the AnLy Xq13.1 breakpoint, and this was subsequently confirmed on an independently derived hybrid line (Grzeschik and Siniscalco 1976). If the translocation breakpoint is actually located within the disease locus, as has been demonstrated in several manifesting female translocation carriers with Duchenne muscular dystrophy, then, barring a significant chromosome rearrangement in the somatic cell lines,

we would suggest that the PGK1 locus is distal to HED (Koenig et al. 1987; Boyd et al., in press). Although the multipoint mapping weakly suggests that HED may be distal to the PGK1 locus, this inference is based on a minimal number of multiply informative recombinants, including one possible recombinant whose carrier status is questionable, and this inference is probably less reliable than the physical mapping data. Recently, several investigators have used panels of X-chromosome deletions and duplications in somatic cell hybrids to physically map many marker loci to the pericentromeric region (Oberle et al. 1986; Arveiler et al. 1987*b*; F. P. M. Cremers, personal communication). One of the somatic cell lines used by each group to delimit an Xq12-q13 breakpoint was originally constructed from the AnLy fibroblast line. Their combined data would place the DXS159 locus proximal to the Xq13.1 breakpoint and would place PGK1 and DXS72 distal to the breakpoint and suggests the following order: DXS159-HED-PGK1-DXS72.

Finally, interfamilial clinical heterogeneity has been observed in X-linked H.E.D., and such variation has also been observed in our study (Freire-Maia and Pinheiro 1984; Clarke et al. 1987*a*). The question has been posed whether this is due to allelic genetic heterogeneity at a single X locus or to multiple X loci. In our 36 families the results of the tests for heterogeneity between HED and eight of the nine marker loci would indicate that H.E.D. is usually produced by a mutation at a single genetic locus and that interfamilial variability is due to different allelic mutations. This information is critical to the clinical application of linkage analysis for both carrier detection and prenatal diagnosis.

Eighty-one percent of females heterozygous for H.E.D. in our study were informative for one of the three marker loci mapped to within 2 cM of the disorder. If one assumes that the AnLy translocation places the disorder between DXS159 and PGK1, then 32% of the carriers have both proximal and distal informative flanking markers within 2 cM of the disorder. It is hoped that further genetic and physical mapping of the disorder will confirm this critical gene order. Linkage analysis of families, with the three closely linked marker loci (DXS159, PGK1, DXS72), combined, if necessary, with analysis of the five additional linked loci, should provide both greatly improved carrier detection for most women at risk for H.E.D. and accurate prenatal diagnosis for many known carriers of the disorder.

Acknowledgments

We would like to thank the investigators listed in table 1 for supplying the DNA probes utilized in this study. We are grateful for the cooperation of the National Foundation for Ectodermal Dysplasia (Mascoutah, IL) for their help in ascertaining the U.S. families. We also appreciate the assistance of the British Paediatric Surveillance Unit; the many genetic centers, physicians, and dentists in the United Kingdom for ascertainment of the British families; and Prof. Albert Schinzel for data on the Swiss family. Thanks to Dr. Robert Koler and Marilyn Jones at Oregon Health Sciences University for assistance in collection and preparation of blood specimens. Finally, we would like to especially thank all of the families whose participation made this study possible. This work was presented in part at the Ninth International Workshop on Human Gene Mapping, held September 6–11, 1987, in Paris and at the American Society of Human Genetics annual meeting, held October 7–10, 1987, in San Diego.

References

Aldridge, J., L. Kunkel, G. Bruns, U. Tantravahi, M. Lalande, T. Brewster, E. Moreau, M. Wilson, W. Bromley, T. Roderick, and S. Latt. 1984. A strategy to reveal high-frequency RFLPs along the human X-chromosome. Am. J. Hum. Genet. 36:546–564.

Arnold, M. L., R. Rauskolb, I. Anton-Lamprecht, A. Schinzel, and W. Schmid. 1984. Prenatal diagnosis of anhidrotic ectodermal dysplasia. Prenat. Diagn. 4:85–98.

Arveiler, B., M. H. Hofker, A. A. B. Bergen, P. Pearson, and J. L. Mandel. 1987*a*. A PstI RFLP detected by probe cpX73 (DXS159) in Xq11-q12. Nucleic Acids Res. 15:5903.

Arveiler, B., I. Oberle, and J. L. Mandel. 1987*b*. Genetic mapping of nine DNA markers in the q11-q12 region of the human X chromosome. Genomics 1:60–66.

Bailit, H. L. 1975. Dental variations among populations. Dent. Clin. North Am. 19:125–140.

Blecher, S. R. 1986. Anhidrosis and absence of sweat glands in mice hemizygous for the Tabby gene: supportive evidence for the hypothesis of homology between Tabby and human anhidrotic (hypohidrotic) ectodermal dysplasia (Christ-Siemens-Touraine) Invest. Dermatol. 87:720–722.

Boyd, Y., E. Munro, S. Holt, D. Cockburn, S. Craig, D. Hunter, and I. Craig. Fine mapping of Xp21 using translocation breakpoints associated with DMD. Cytogenet. Cell Genet. (in press).

Bruns, G., J. Aldridge, L. Kunkel, U. Trantravahi, M. Lalande, T. Dryja, and S. A. Latt. 1984. Molecular analysis of the human X-chromosome. Cytogenet. Cell Genet. 37:428–429.

Clarke, A., D. I. M. Phillips, R. Brown, and P. S. Harper. 1987*a*. Clinical aspects of X-linked hypohidrotic ectodermal dysplasia. Arch. Dis. Child. 62:989–996.

Clarke, A., M. Sarfarazi, N. S. T. Thomas, K. Roberts, and P. S. Harper. 1987*b*. X-linked hypohidrotic ectodermal dysplasia: DNA probe linkage analysis and gene localization. Hum. Genet. 75:378–380.

Cohen, M. M., C. C. Lin, V. Sybert, and E. J. Orecchio. 1972. Two human X-autosome translocations identified by autoradiography and fluorescence. Am. J. Hum. Genet. 24:583–597.

Conneally, P. M., J. H. Edwards, K. K. Kidd, J. M. Lalouel, N. E. Morton, J. Ott, and R. White. 1985. Report of the Committee on Methods of Linkage Analysis and Reporting. Cytogenet. Cell Genet. 40:356–359.

Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132:6–13.

Freire-Maia, N., and M. Pinheiro. 1984. Christ-Siemens-Touraine's (CST) syndrome. Pp. 33–36 *in* Ectodermal dysplasias: a clinical and genetic study. Alan R. Liss, New York.

Frias, J. L., and D. W. Smith. 1968. Diminished sweat pores in hypohidrotic ectodermal dysplasia: a new method for assessment. J. Pediatr. 72:606–610.

Gerald, P. S., and J. A. Brown. 1974. Report of the Committee on the Genetic Constitution of the X Chromosome. Cytogenet. Cell Genet. 13:29–34.

Grzeschik, K. H., and M. Siniscalco. 1976. Identification of a de novo chromosome rearrangement in a man-mouse hybrid clone and its bearing on the cytological map of the human X chromosome. Cytogenet. Cell Genet. 16:149–156.

Happle, R., and P. J. Frosch. 1985. Manifestation of the lines of Blaschko in women heterozygous for X-linked hypohidrotic ectodermal dysplasia. Clin. Genet. 27:468–471.

Hodgson, S. V., M. E. Robertson, C. N. Fear, J. Goodship, S. Malcolm, B. Jay, M. Bobrow, and M. E. Pembrey. 1987. Prenatal diagnosis of X-linked choroideremia with mental retardation, associated with a cytologically detectable X-chromosome deletion. Hum. Genet. 75:286–290.

Hutz, M. H., A. M. Michelson, S. E. Antonarakis, S. H. Orkin, and H. H. Kazazian. 1984. Restriction site polymorphism in the phosphoglycerate kinase gene on the X chromosome. Hum. Genet. 66:217–219.

Keith, D. H., J. Singer-Sam, and A. D. Riggs. 1986. Active X chromosome DNA is unmethylated at eight CCGG sites clustered in a guanine-plus-cytosine-rich island at the 5′ end of the gene for phosphoglycerate kinase. Mol. Cell. Biol. 6:4122–4125.

Kleinebrecht, J., K. H. Degenhardt, A. Grubisic, E. Gunther, and J. Svejcar. 1981. Sweat pore counts in ectodermal dysplasias. Hum. Genet. 57:437–439.

Koenig, M., E. P. Hoffman, C. J. Bertelson, A. P. Monaco, C. Feener, and L. M. Kunkel. 1987. Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell 50:509–517.

Koenig, M., J. P. Moisan, R. Heilig, and J. L. Mandel. 1985. Homologies between X and Y chromosomes detected by DNA probes: localisation and evolution. Nucleic Acids Res. 13:5485–5501.

Kolvraa, S., T. Kruse, P. K. A. Jensen, K. H. Linde, S. R. Vestergaard, and L. Bolund. 1986. Close linkage between X-linked ectodermal dysplasia and a cloned DNA sequence detecting a two allele restriction fragment length polymorphism in the region Xp11-q12. Hum. Genet. 74:284–287.

Kunkel, L. M., K. D. Smith, S. H. Boyer, D. S. Borgaonker, S. S. Wachtel, O. J. Miller, W. R. Breg, H. W. Jones, and J. M. Rary. 1977. Analysis of human Y-chromosome-specific reiterated DNA in chromosome variants. Proc. Natl. Acad. Sci. USA 74:1245–1249.

Lathrop, G. M., J. M. Lalouel, C. Julier, and J. Ott. 1984. Strategies for multilocus linkage analysis in humans. Proc. Natl. Acad. Sci. USA 81:3443–3446.

Lesko, J. G., R. A. Lewis, and R. L. Nussbaum. 1987. Multipoint linkage analysis of loci in the proximal long arm of the human X chromosome: application to mapping the choroideremia locus. Am. J. Hum. Genet. 40:303–311.

MacDermot, K. D., R. M. Winter, and S. Malcolm. 1986. Gene localisation of X-linked hypohidrotic ectodermal dysplasia (C-S-T syndrome). Hum. Genet. 74:172–173.

Nakata, M., H. Koshiba, K. Eto, and W. E. Nance. 1980. A genetic study of anodontia in X-linked hypohidrotic ectodermal dysplasia. Am. J. Hum. Genet. 32:908–919.

Oberle, I., G. Camerino, C. Kloepfer, J. P. Moisan, K. H. Grzeschik, B. Hellkuhl, M. C. Hors-Cayla, N. Van Cong, D. Weil, and J. L. Mandel. 1986. Characterization of a set of X-linked sequences and of a panel of somatic cell hybrids useful for the regional mapping of the human X chromosome. Hum. Genet. 72:43–49.

Ohno, S. 1973. Ancient linkage groups and frozen accidents. Nature 244:259–262.

Ott, J. 1983. Linkage analysis and family classification under heterogeneity. Ann. Hum. Genet. 47:311–320.

———. 1985. Homogeneity tests. Pp. 111–119 *in* Analysis of human genetic linkage. Johns Hopkins University Press, Baltimore.

Page, D., B. De Martinville, A. Barker, A. Wyman, R. White, U. Francke, and D. Botstein. 1982. Single-copy sequence hybridizes to polymorphic and homologous loci on human X and Y chromosomes. Proc. Natl. Acad. Sci. USA 79:5352–5356.

Pinheiro, M., M. T. Ideriha, E. A. Chautrad-Freire-Maia, N. Freire-Maia, and S. L. Primo-Parmo. 1981. Christ-Siemens-Touraine syndrome: investigations on two large

Brazilian kindreds with a new estimate of the manifestation rate among carriers. Hum. Genet. 57:428–431.

Riddell, D. C., H. S. Wang, J. Beckett, A. Chan, J. J. A. Holden, L. M. Mulligan, M. A. Phillips, N. E. Simpson, K. Wrogemann, J. L. Hamerton, and B. N. White. 1986. Regional localization of 18 human X-linked sequences. Cytogenet. Cell Genet. 42:123–128.

Schmeckpeper, B., J. Davis, H. Willard, and K. Smith. 1985. An anonymous single-copy X chromosome RFLP for DXS72 from Xq13-q22. Nucleic Acids Res. 13:5724.

Schwartz, M., H. M. Yang, E. Niebuhr, T. Rosenberg, and D. C. Page. Deletion mapping of the choroideremia locus (TCD) and flanking polymorphic X chromosomal loci. Cytogenet. Cell Genet. (in press).

Searle, A. G., J. Peters, M. F. Lyon, E. P. Evans, J. H. Edwards, and V. J. Buckle. 1987. Chromosome maps of man and mouse, III. Genomics 1:3–18.

Shows, T. B., and J. A. Brown. 1974. An (Xq+;9p−) translocation suggests the assignment of G6PD, HPRT, and PGK to the long arm of the X chromosome in somatic cell hybrids. Cytogenet. Cell Genet. 13:146–149.

———. 1975. Human X-linked genes regionally mapped utilizing X-autosome translocations and somatic cell hybrids. Proc. Natl. Acad. Sci. USA 72:2125–2129.

Stevenson, A. C., and C. B. Kerr. 1967. On the distributions of frequencies of mutation to genes determining harmful traits in man. Mutat. Res. 4:339–352.

Vogelstein, B., E. R. Fearon, S. R. Hamilton, A. C. Preisinger, H. F. Willard, A. M. Michelson, A. D. Riggs, and S. H. Orkin. 1987. Clonal analysis using recombinant DNA probes from the X-chromosome. Cancer Res. 47:4806–4813.

Wieacker, P., K. E. Davies, H. J. Cooke, P. L. Pearson, R. Williamson, S. Bhattacharaya, J. Zimmer, and H.-H. Ropers. 1984. Toward a complete linkage map of the human X chromosome: regional assignment of 16 cloned single-copy DNA sequences employing a panel of somatic cell hybrids. Am. J. Hum. Genet. 36:265–276.

Yang, H. M., E. Niebuhr, T. Rosenberg, and M. Schwartz. An X chromosome deletion panel as a tool for subregional assignment of X chromosomal probes and disease loci. Cytogenet. Cell Genet. (in press).

Zonana J., S. H. Roberts, N. S. T. Thomas, and P. S. Harper. Recognition and re-analysis of a cell line from a manifesting female with X-linked hypohidrotic ectodermal dysplasia and an X; autosome balanced translocation. J. Med. Genet. (in press).

GENOMICS 16, 572–579 (1993)

Exhibit C

Novel Triplet Repeat Containing Genes in Human Brain: Cloning, Expression, and Length Polymorphisms

SHI-HUA LI,* MELVIN G. MCINNIS,*,† RUSSELL L. MARGOLIS,*
STYLIANOS E. ANTONARAKIS,† AND CHRISTOPHER A. ROSS,*,[1]

*Laboratory of Molecular Neurobiology, Departments of Psychiatry and Neuroscience, Ross 618, Johns Hopkins University, School of Medicine, 720 Rutland Avenue, Baltimore, Maryland 21205–2196; and †Center for Medical Genetics and Department of Pediatrics and Medicine, Johns Hopkins University, School of Medicine, 600 N. Wolfe Street, Baltimore, Maryland 21287–3914

Received December 7, 1992; revised January 28, 1993

Human genes containing triplet repeats may markedly expand in length and cause neuropsychiatric disease, explaining the phenomenon of anticipation (increasing severity or earlier age of onset in successive generations in a pedigree). To identify novel genes with triplet repeats, we screened a human brain cDNA library with oligonucleotide probes containing CTG or CCG triplet repeats. Fourteen of 40 clones encoded novel human genes, and 8 of these inserts have been sequenced on both strands. All contain repeats, and 5 of the 8 have 9 or more consecutive perfect repeats. All are expressed in brain. Chromosomal assignments reveal a distribution of these genes on multiple autosomes and the X-chromosome. Further, the repeat length in two of the genes is highly polymorphic, making them valuable index linkage markers. We predict that many triplet repeat-containing genes exist; screening with the CTG probe suggests approximately 50–100 genes containing this type of repeat are expressed in the human brain. Since additional disorders, such as Huntington's disease, bipolar affective disorder, and possibly others, show features of anticipation, we suggest that these novel human genes with triplet repeats are candidates for causing neuropsychiatric diseases.

INTRODUCTION

Three neuropsychiatric disorders are caused by expansion of unstable triplet repeats in human genes (Caskey *et al.*, 1992; Davies, 1992; Richards and Sutherland, 1992). Myotonic dystrophy is an autosomal dominant disorder characterized primarily by myotonia and progressive muscle weakness. Expansion of CTG triplet repeats in a gene called myotonin protein kinase underlies expression of the disease (Mahadevan *et al.*, 1992; Brook *et al.*, 1992; Harley *et al.*, 1992a). Normals have a range of 5–30 repeats, whereas patients with myotonic dystrophy often have many hundreds of repeats. The degree of expansion correlates with severity of the disease and increases in successive generations (Tsilfidis *et al.*, 1992). This observation explains the previously puzzling feature of the inheritance of this disorder, termed "anticipation"—decreasing age of onset and increasing severity of the disorder in successive generations within a pedigree (Howeler *et al.*, 1989; Dyken and Harper, 1973; Harley *et al.*, 1992b). Fragile X syndrome, the most common defined form of inherited mental retardation, is also caused by expansion of an unstable triplet repeat, in this case a CGG repeat within a gene known as FMR-1 (Fu *et al.*, 1991; Verkerk *et al.*, 1991; Kremer *et al.*, 1991). This disorder also has unusual features of inheritance consistent with anticipation, termed the "Sherman paradox" (Caskey *et al.*, 1992; Sherman *et al.*, 1985). The degree of expansion appears to correlate with the extent of penetrance of the phenotype (Caskey *et al.*, 1992; Yu *et al.*, 1992). Finally, Kennedy disease, or X-linked spinal and bulbar muscular atrophy, is caused by an expanding CAG repeat, coding for polyglutamine, within the first exon of the androgen receptor gene (La Spada *et al.*, 1991; Tilley *et al.*, 1989).

In addition to the genes identified as causing neuropsychiatric illness, other genes with triplet repeats are already known. In drosophila and mouse, genes with CAG or CAA repeats encoding the amino acid glutamine are called *OPA* repeats and are frequently found in homeodomain genes or other DNA binding proteins (Wharton *et al.*, 1985; Duboule *et al.*, 1987). *Pen* repeats are GGN repeats often coding for long glycine stretches (Haynes *et al.*, 1987).

As discussed below, other neuropsychiatric diseases show features of anticipation, suggesting that they could be caused by expansion of triplet repeats in currently unknown genes. We therefore sought to determine whether additional genes with triplet repeats are expressed in the human brain.

MATERIALS AND METHODS

cDNA cloning. To identify triplet repeat containing genes expressed in brain we screened a cerebral cortex cDNA library in lambda

[1] To whom correspondence and reprint requests should be addressed. Telephone: (410) 614-0011. Fax: (410) 614-0013.

0888-7543/93 $5.00

ZAP (Stratagene) with oligonucleotides $(CTG)_{10}$ or $(CCG)_{10}$ using standard techniques (Sambrook *et al.*, 1989). Oligonucleotides were labeled with [γ-^{32}P]ATP and polynucleotide kinase to a specific activity of 3000–4000 Ci/mmol. The library was plated at a density of approximately 50,000 plaques per 150-mm plate. Baked duplicate filter lifts, washed with 2× SSC (0.3 *M* NaCl, 0.03 *M* sodium citrate, pH 7) for 30 min, and prehybridized with a hybridization buffer containing 5× SSPE (0.9 *M* NaCl, 5 m*M* EDTA, and 50 m*M* $NaH_2PO_4 \cdot H_2O$), 5× Denhardt's solution (0.1% Ficoll 400, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin), 10 μg/ml of denatured salmon sperm DNA and 50% formamide at 42°C for at least 2 h were hybridized at 42°C overnight in the same buffer containing labeled CGG or CTG oligo probes (1–3 × 10^6 cpm/ml). Hybridization with CCG probe was carried out at 48°C overnight. Wash for the CTG oligo was at 60°C in 1× SSC/0.5% SDS (1× SSC= 0.15 *M* NaCl, 0.015 *M* sodium citrate, pH 7) for 20 min ×3; wash for the CGG oligo was at 65°C in 1× SSC/0.5% SDS for 20 min ×3. Thirty-eight positives were selected and plaque purified (14 from the CGG oligonucleotide screening and 24 from the CTG oligonucleotide screening). The plasmids (pBluescript SK) were excised using a commercial *in vivo* excision procedure (Stratagene) and the ends of the clones were sequenced using T3 and T7 primers.

Both ends of nucleotide sequence from the clones were used for homology search against GenBank (release 71) using the FastA program (GCG package, version 7, on a VAX 8530 computer). The inserts of clones with ends containing novel human sequences were then sequenced in their entirety on both strands using an ABI automated sequencing apparatus by Roxann Ingersoll in the Johns Hopkins Genetics Core Facility. Nucleotide sequences of the inserts were analyzed for open reading frames and predicted peptide sequences for hydrophobicity using the MacVector program with a Macintosh IIci computer. The starting codons in the putative peptides were assigned as the first methionine after stop codons in all three reading frames. In the clones with a single long open reading frame the peptide sequence was also used for a GenBank homology search using the TFastA program.

Northern blot analysis. To determine whether or not these genes were expressed in human brain, we conducted Northern analysis. Total RNA was extracted from human frontal cerebral cortex, human cerebellar cortex, and total mouse brain, using the cesium chloride ultracentrifugation method (Davis *et al.*, 1986). Total RNA was denatured by heating at 65°C for 10 min with 50% formamide, separated electrophoretically on a denaturing formaldehyde agarose gel, transferred to nitrocellulose, and probed with 45 mer oligonucleotide probes designed to hybridize to a region of the gene excluding the repeat. In cases in which the clone had a poly(A) tail, we made a probe only in the antisense orientation. In cases in which the direction was not established by a poly(A) tail, we made probes in both the presumed sense and the presumed antisense orientation based on open reading frame analysis. In all cases, Northern hybridization yielded a signal with the antisense probe but not with the sense probe. Blots were prehybridized with a buffer containing 4× SSC, 0.01 *M* Tris-HCl, pH 7.5, 2× Denhardt's, 200 μg/ml denatured salmon sperm DNA, and 40% formamide at 42°C for at least 3 h. Oligonucleotides were labeled using [γ-^{32}P]ATP and polynucleotide kinase and hybridized overnight with the blots in the same buffer plus 10% dextran sulfate and 2–3 × 10^6 cpm/ml labeled probes at 42°C. The specific activity of the oligo probes was approximately 3600 Ci/mmole. Blots were washed first at room temperature in 2× SSC/0.1% SDS for 15 min ×3 and then, for higher stringency, at 55 to 60°C in 0.2× SSC/0.1–0.5% SDS for 15 min ×3. Exposures at −70°C with an intensifying screen ranged in length from overnight to 4 days.

Polymorphism analysis. To detect the presence of polymorphisms in the lengths of the repeats we used PCR analysis of human genomic DNA from parents in the CEPH families (Dausset *et al.*, 1990). Primers were chosen using the program *Oligo* (National Bioscience Inc.) to yield a product of 100 to 200 bases spanning the repeat (Table 3). PCR was performed using modified techniques of Sambrook *et al.* (1989). Briefly, after an initial denaturation step of 96°C for 2 min, the cycle was 55–67°C annealing for 1 min (Table 3), 72°C extension for 1 min, and 96°C denaturation for 1 min for a total of 30–32 cycles with a final extension step for 10 min.

The amplification of a specific band was first optimized using nonradioactive PCR and then analysis was carried out with one of the two primers labeled with [γ-^{32}P]ATP and polynucleotide kinase. The reaction volume was 25 μl, containing 200 ng of template DNA, 0.42 μM of each primer, 200 μM DNTPs, 1.5 m*M* magnesium chloride, and 0.5 units of *Taq* polymerase (Perkin-Elmer Cetus). All experiments included a negative control using the same conditions and water as a template and a positive control using 10 pg to 10 ng of the cloned cDNA as a template. The individuals routinely studied were parents from 12 randomly chosen pedigrees of the CEPH families (48 chromosomes). For nonradioactive PCR, the products were electrophoretically separated on a 2 to 3% agarose gel. For radioactive PCR experiments, the products were separated on a 6% denaturing polyacrylamide gel and exposed to film for durations ranging from 3 h to overnight.

Chromosomal localization. For chromosomal localization of the clones we used a panel of monochromosomal human–rodent hybrid cell lines as templates for PCR. Each cell line contains a single human chromosome. A negative control with water as the template and a positive control with the cloned plasmid DNA as the template were included in each experiment. All reaction conditions were optimized such that a product formed when human genomic DNA was used as a template but no product (or product of a different size) was formed when either hamster or mouse genomic DNA at the same concentration served as the template. The PCR conditions were the same as described in polymorphism analysis (with 100 to 250 ng of DNA for each template). The products were electrophoretically separated on an agarose or acrylamide gel as described above. Each chromosomal localization experiment was repeated at least twice.

RESULTS

Cloning of Novel Human Genes with Triplet Repeats

To identify genes with triplet repeats expressed in the human brain, we screened a human frontal cortex cDNA library with oligonucleotide probes containing 10 repeats of either CTG or CGG. There were approximately 100 positives per 150-mm plate (of about 50,000 plaques) in the initial screen for the CTG probe and many more positives for the CGG probe.

To form a more accurate estimate of the number of clones containing triplet repeats, filters were lifted from lower density plates (5000–6000 plaques). Of these, 0.28% were positive with the CTG probe and 4.7% were positive with the CGG probe. At least four-fifths of the latter, however, encoded 28S ribosomal RNA, as demonstrated by hybridization with a 28S ribosomal RNA oligonucleotide pool probe.

Of our initial screen, 40 positives were selected for plaque purification. Two did not show a strong signal on rescreen and were discarded. The sequences of the remaining clones indicated that 12 coded for 28S ribosomal RNA, which is known to have CGG repeats (Gonzalez *et al.*, 1985). Nine were previously known genes or were human homologues of known mammalian genes (Table 1) including human apoferritin, sodium potassium ATPase, and calcium-dependent protease (Costanzo *et al.*, 1984; Lane *et al.*, 1989; Ohno *et al.*, 1986). The first clone sequenced, CTG-A2, was myotonin protein kinase, with 13 CTG repeats. Three of the clones could not be readily sequenced. Fourteen were novel hu-

TABLE 2

Characterization of the Clones

Clones	mRNA kb (exposure)	Chromosomal localization	Number of alelles	Percentage of heterozygosity
CCG-A3	4.5 (2 days)	5, 6	1	0
CTG-A4	5.0 (3 days)	13	2	20
CTG-B1	5.5 (4 days)	1	2	50
CCG-B7	2.3 (2 days)	X	ND	ND
CTG-B33	1.2 (2 days)	3, (12)	7	70
CTG-B37	2.4 (3 days)	12	13	85
CTG-B43	2.0 (4 days)	16	1	0
CTG-B45d	9.0 (3 days)	10, (20)	1	0

Note. ND, not done.

Chromosomal Localization

For the chromosomal assignment of these genes, we used DNA from rodent–human hybrid cell lines containing a single human chromosome (Warburton *et al.*, 1990; Theune *et al.*, 1991) as PCR templates (see Table 2 and Fig. 3). Each clone was assigned to a different autosome, except for the assignment of CCG-B7 to the X chromosome. Three genes yielded products with DNA from two chromosomes by several different primer pairs, suggesting the existence of pseudogenes or closely related genes (on chromosomes 5 and 6 for CCG-A3, on chromosomes 3 and 12 for CTG-B33, and on chromosomes 10 and 20 for CTG-B45d). Radioactive PCR with DNA templates from monochromosomal rodent–human hybrid cell lines showed that the products of chromosomes 3 and 10 are the same size as of the products of clones CTG-B33 and CTG-B45d, respectively. The products of chromosomes 12 and 20, respectively, were of different size and thus are shown in parentheses in Table 2. The products of both chromosome 5 and chromosome 6 showed the same size as clone CCG-A3 and thus both are shown without parentheses (Table 2).

Polymorphisms in Lengths of the Triplet Repeats

To analyze the lengths of the triplet repeats in a sampling (48 chromosomes) of the normal human population, we used PCR primers spanning the repeats with templates from a selection of parents in the CEPH pedigrees (Table 2 and Fig. 4). Two of the repeats were very highly polymorphic.

CTG-B37 had 13 different alleles with 85% observed heterozygosity in an analysis of 24 CEPH parents (Fig. 4a). The copy numbers of the CAG repeat ranged from 9 to 23.

Clone CTG-B33 had 7 alleles on chromosome 3; about 70% of individuals were heterozygous. In addition, the cross-reacting band on chromosome 12 also revealed heterozygosity, with 3 or 4 alleles and 20% heterozygosity (Fig. 4b).

Clone CTG-B1 showed about 50% heterozygosity with 2 alleles (not shown). Clone CTG-A4 showed only slight heterozygosity (20%) with 2 alleles and clones CCG-A3, CTG-B45d and CTG-B43a were not polymorphic (not shown).

To confirm the inheritance of the different alleles of clone CTG-B37, we conducted PCR analysis in the CEPH family No. 1347 (Fig. 5). The alleles were inherited in a Mendelian fashion through all three generations. Preliminary linkage analysis is consistent with the assignment to chromosome 12.

DISCUSSION

In this study we have identified 14 novel human genes containing triplet repeats and analyzed 8 of them. We find that these genes are widely dispersed among the human chromosomes. They are expressed as scarce to moderately abundant messages in human cerebral and cerebellar cortices. In several cases the lengths of the

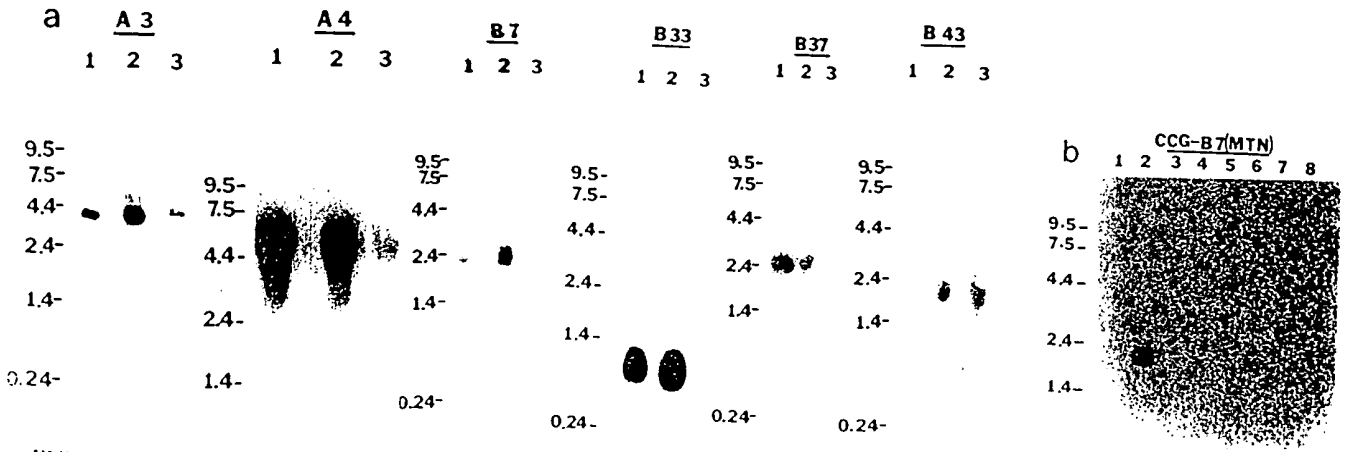


FIG. 2. Expression of the triplet repeat containing genes in human brain by RNA expression analysis. Northern blots of total RNA (20 μg per lane) from human frontal cerebral cortex (lanes 1), human cerebellar cortex (lanes 2), and mouse brain (lanes 3), probed with oligonucleotide probes for each of the indicated clones. Most of the clones are expressed in human brain at low to moderate abundance (a). The exposure times are indicated in Table 2. (b) (CCG-B7 MTB) is a Northern blot of poly(A$^+$) RNA from human heart (lane 1), brain (lane 2), placenta (lane 3), lung (lane 4), liver (lane 5), skeletal muscle (lane 6), kidney (lane 7), and pancreas (lane 8), which was hybridized with an oligo probe from CCG-B7 and exposed for 48 h.

TABLE 3

Primers for PCR

Clone	Primers	Size of products (bp)	Annealing temperature (°C)
CCG-A3	5′(296)TCGAGACCAGAACGGCCTTTCCAA 5′(547)TATTACTGCGACCGCCCCAGCTAC	250	59
CTG-A4	5′(302)CTGATCACTTGTGGTTCTGCGCCG 5′(485)AGGACTCCCCCATGCTCGCCAG	183	67
CTG-B1	5′(270)AGTTCCGCCAGCCCTTGAGAG 5′(381)GGGAGCTGCCTGCCTCTTGAT	111	62
CCG-B7	5′(153)CTGGCTGTTGGAGTCTGGGGC 5′(316)GCATCCATGGGCTACCACGAC	163	63
CTG-B33	5′(425)CAAAAAAGCACCTGGTACTAA 5′(541)GGGCTGGAGCCTTTTTACTCGC	116	59
CTG-B37	5′(699)CACCAGTCTCAACACATCACCATC 5′(840)CCTCCAGTGGGTGGGGAAATGCTC	141	62
CTG-B43	5′(865)CCCACCATCTTCCCGCTGC 5′(980)GCAGCTGGGCAGTCTGGGC	115	55
CTG-B45d	5′(463)CTCCAGGAAGAGGAAGAAAGC 5′(619)CTGTGGCTCAACCCTCATTTC	156	65

Note. PCR primers for chromosomal assignment and length polymorphism. Primers for clones CTG-B7 and CTG-B43 do not span the repeats. PCR primers for other clones amplify a DNA fragment across the repeats. The numbers in parenthesis are the base number of PCR primers' 5′ position.

repeats are highly polymorphic, making them valuable as PCR typeable linkage markers.

Previous studies have found that triplet repeats are enriched relative to expected rates in genomic DNA. AAC, CCG (=CGG), AGC (=CTG), and AAT are the most common motifs found in sequence data bases (Tautz and Renz, 1984; Beckmann and Weber, 1992). One previous study has used oligonucleotide probes to identify several new CAC repeats in genomic DNA; however, no mRNA transcripts were found in various human tissues (Zischler *et al.*, 1992). Another group has used PCR to identify simple sequence repeats in human genomic DNA (Edwards *et al.*, 1991). Like minisatellites, triplet repeats have been found to be widely dispersed in human genomic DNA (Zischler *et al.*, 1992; Edwards *et al.*, 1991; Tautz, 1989; Wong *et al.*, 1987). We now find that genes containing triplet repeats are also present on many different chromosomes.

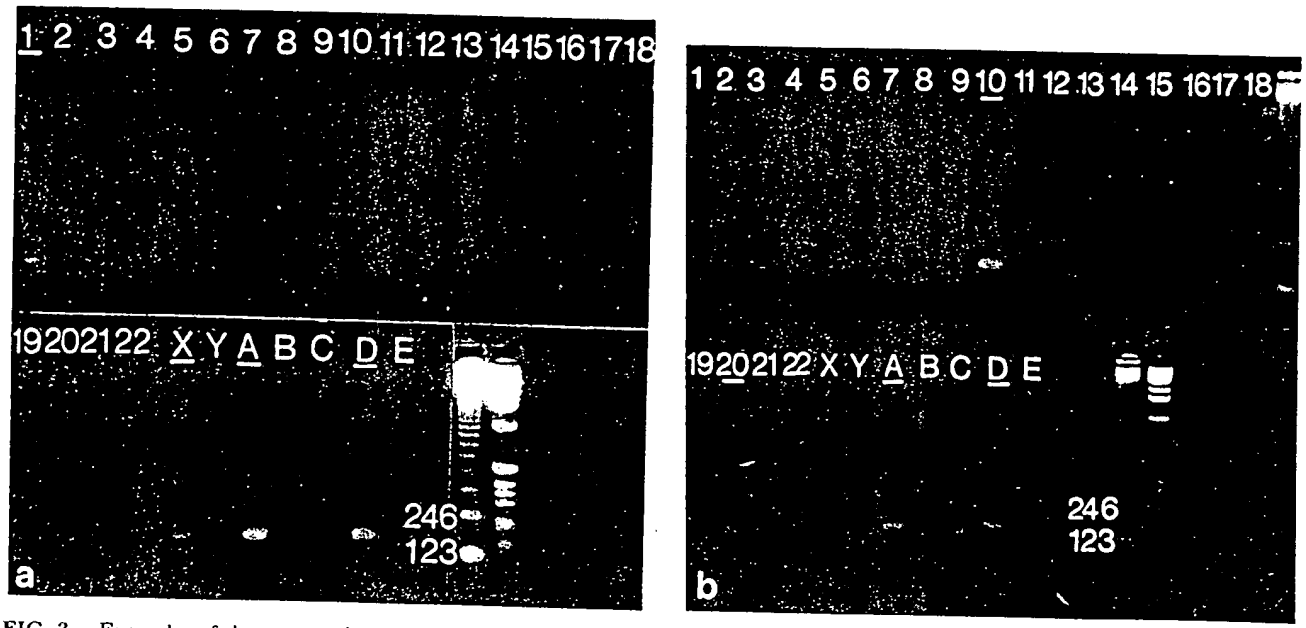


FIG. 3. Examples of chromosomal assignments. Primers spanning the repeats or spanning a region adjacent to the repeats were used for a PCR, with single human chromosomal human–rodent hybrid cell lines. (a) Clone CCG-B7 is assigned to the X chromosome; lanes 1–22, X, and Y are the PCR products of human–rodent hybrid cell DNA as indicated except that lane 1 has DNA from both chromosomes 1 and X; lane A, human genomic DNA; lane B, mouse genomic DNA; lane C, Chinese hamster genomic DNA; lane D, clone CCG-B7 plasmid DNA; lane E, water as template; 246 and 123 indicate size markers. (b) Clone CTG-B45d is assigned to chromosome 10. (There is also a band at chromosome 20, but radioactive PCR indicates that the alleles corresponding to the clone are on chromosome 10 as shown in Fig 4 for CTG-B33.) The lanes are the same as in (a).

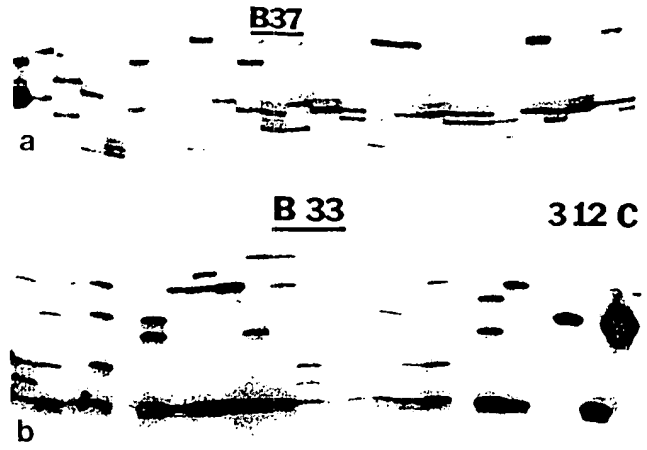


FIG. 4. Repeat lengths of two triplet repeats are highly polymorphic. Radioactive PCR was performed using the primers spanning the repeats in CTG-B33 or CTG-B37 with DNA from parents in the CEPH pedigrees (Nos. 1413, 1416, 1362, 1421, 66, 1418, 12, 884, 13291, 1350, 1331, 1332). Clone CTG-B37 has 13 alleles and is highly polymorphic (a). For clone CTG-B33 (b) the highly polymorphic allele is on chromosome 3; lane 3 is the amplification of chromosome 3 human–rodent hybrid DNA; lane 12 is the amplification of chromosome 12 human–rodent hybrid DNA; and lane C is the amplification of clone CTG-B33 plasmid with radiolabeled PCR primers. CTG-B33 has approximately 7 alleles.

A recently published paper presents data consistent with those of the current study (Riggins *et al.*, 1992). The lengths of the triplet repeats of several genes in the database were found to be highly polymorphic. In addition, Riggins *et al.* (1992) screened a number of cDNA libraries with $(CAG)_{10}$ and $(CGG)_{10}$, finding a similar abundance of positives in fetal brain as we find in adult brain, but finding 5- to 10-fold fewer positives in libraries from various cell lines and muscle and testis. The lengths of the repeats of these clones were shorter, and only slightly polymorphic, perhaps because the screening was done at a lower stringency than in the current study.

The present study establishes that the novel genes with triplet repeats are expressed as mRNAs in human brain. Based on the results of screening reported here we estimate that on the order of 50–100 novel genes containing CTG or CAG repeats are expressed in human brain (assuming that 50,000–100,000 genes are expressed in brain, that the representation of the triplet repeat-containing genes parallels that of all genes in the library, and that about one-sixth to one-third of positives are independent genes). The number of CCG or CGG repeats is harder to estimate, since so many of our positives were 28S ribosomal RNA, but could be greater.

In 7 of the 8 clones in which we sequenced the full insert, the triplet repeat was contained within an open reading frame. Thus, it is likely that the repeats are translated into amino acids, including glutamine, glycine, proline, and serine (Fig 1). Hydrophobicity analysis suggests that all except one (CCG-B7) of the proteins are soluble. This is in contrast to the repeats of Riggins *et al.* (1992), which were mostly in the 5′ untranslated region. This could be because most of our novel sequences were CAG repeats, whereas most of the were CGG.

The high degree of heterozygosity in the lengths these triplet repeats suggest that they, like dinucleoti repeats or tetranucleotide repeats, may be useful linka markers in the human genome project (Tautz and Re 1984; Tautz, 1989; Warren *et al.*, 1992; Edwards *et* 1992). For CAA repeats the longest run of uninterrupt repeats has been found to be the best predictor of infc mativeness (Weber, 1990). Perfect repeat sequences about 30 nucleotides or greater are likely to have a rel tively high heterozygosity. Similarly, we find for t CAG repeats that the genes with the longest number uninterrupted repeats in our initial clones (CTG-B with 13 and CTG-B37 with 15) had the highest degree polymorphism. The experience with the three diseas so far known to be caused by expansion of triplet repea suggests that repeats that are highly polymorphic in t normal population may be more likely to expand wit pathogenic consequences.

The function of these repeats, if any, is only partial understood. Simple sequences in genomic DNA ma have a role in recombination or in the insertion of tran posable elements (Ivanova *et al.*, 1984; Epplen, 1988 The function of repeated single amino acids in protein also uncertain, although their presence in a number transcription factors suggests they could in some case have a role in protein DNA interaction. *Pen* repeat se quences, isolated in drosophila, encode polyglycine an are present in a homologue of the rat helix destabilizin protein (Haynes *et al.*, 1987). One of our clones (CCG A3) showed homology to this protein. A proline-ric transcriptional activation zone has been describe (Mermod *et al.*, 1988). *Opa* repeat containing proteins i drosophila and mouse, which have polyglutamine tract (Wharton *et al.*, 1985; Duboule *et al.*, 1987), often cod for transcription factors including homeodomain pro teins. Human TATA binding protein has 38 consecutiv glutamine residues in its amino terminal domain (Kao *e al.*, 1990). Sp1 has a glutamine-rich region with shorte

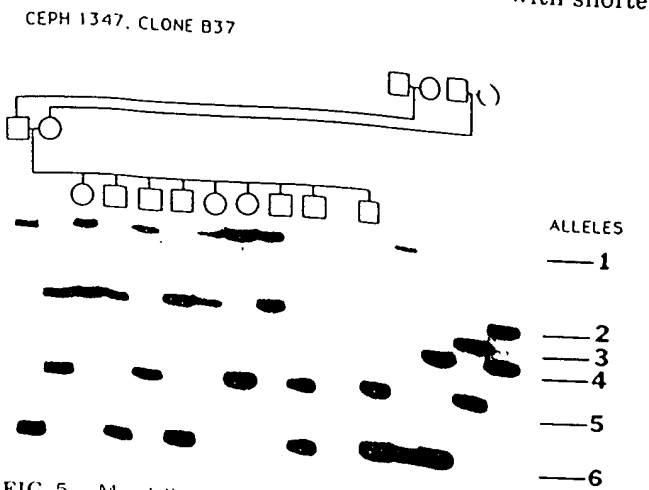


FIG. 5. Mendelian inheritance of the repeat length of clone CTG B37. The pedigree of CEPH family No. 1347 is shown above. The PCR products showed six different alleles in this family.

glutamine repeats, important for transcription (Kadonaga *et al.*, 1988; Courey and Tijan, 1988). Thus some of our clones with glutamine repeats may code for transcription factors. Other repeats may be involved in protein–protein interaction. For instance an 11-proline repeat in human calcineurin A has been proposed to be involved in binding to calmodulin (Guerini and Klee, 1989).

Disorders other than those described in the introduction may have patterns of inheritance consistent with anticipation. In Huntington's disease, offspring of affected males have younger ages of onset than their parents (Ridley *et al.*, 1988). The presence of individuals with "schizophrenia spectrum" disorders in families of probands with schizophrenia (Kendler *et al.*, 1985), suggests variable penetrance, and the risk of developing schizophrenia for children of probands with schizophrenia is more than twice that for parents (Gottesman, 1991). Our recent study of pedigrees with unilineal transmission of bipolar affective illness suggests that disease severity increases and age of onset decreases in transmission from one generation to the next (Depaulo *et al.*, 1992; McInnis *et al.*, 1992).

Our study shows that screening cDNA libraries with oligonucleotide probes can be used to identify novel genes with triplet repeats. Some of these genes may code for transcriptional regulators or other nucleotide binding proteins. Based on our experience many are likely to be highly polymorphic, making these genes valuable linkage markers for studies of the human genome. Finally, as we suggest above, additional neuropsychiatric disorders appear to have features of anticipation consistent with causation by expansion of triplet repeats. Therefore, we suggest that novel human genes with triplet repeats may be candidate genes for neuropsychiatric disorders.

ACKNOWLEDGMENTS

This study was supported by a Pew Scholars Award in the Biomedical Sciences to C.A.R. The authors thank Audrey Bentley for technical assistance.

Note added in proof. The recent discovery that Huntington disease is caused by expansion of a triplet repeat (H.D. Collaborative Group, *Cell*, in press) is consistent with the hypotheses presented here and in our recent review (Ross *et al.*, *Trends Neurosci.*, in press).

REFERENCES

Baron, M., Gruen, R., Rainer, J. D., Kane, J., Asnics, L., and Lord, S. (1985). A family study of schizophrenic and normal control probands: Implications for the spectrum concept of schizophrenia. *Am. J. Psychiatry* **142**: 447–455.

Beckmann, J. S., and Weber, J. L., (1992). Survey of human and rat microsatellites. *Genomics* **12**: 627–631.

Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., Hunter, K., Stanton, V. P., Thirion, J. P., and Hudson, T. (1992). Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. *Cell* **69**: 784–789.

Caskey, C. T., Pizzuti, A., Fu, Y-H., Fenwick, R. G., and Nelson, D. L. (1992) Triplet repeat mutations in human disease. *Science* **256**, 784–789.

Costanzo, F., Santoro, C., Colantuoni, V., Bensi, G., Raugei, G., Romano, V., and Cortese, R. (1984). Cloning and sequencing of a full length cDNA coding for a human apoferritin H chain: Evidence for a multigene family. *EMBO J.* **3**: 23–27.

Courey, A. J., and Tjian, R. (1988). Analysis of Sp 1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. *Cell* **55**: 887–898.

Dausset, J., Cann, H, Cohen, D., Lathrop, M., Lalouel, J. M., and White, R. (1990). Centre d'Etude du Polymorphisme humain (CEPH): Collaborative genetic mapping of the human genom. *Genomics* **6**: 575–577.

Davies, K. (1992). The cost of instability. *Nature* **356**: 15.

Davis, L. G., Dibner, M. D., and Battey, J. F. (1986). "Basic Methods in Molecular Biology," Elsevier, New York.

DePaulo, Jr., McInnis, M. G., McMahon, F. J., Simpson, S. G., and Ross, C. A. (1992). Anticipation in biopolar affective disorder. *Am. J. Hum. Genet.* **51**, a44.

Duboule, D., Maenlin, M., Galliot, B., and Mohier, E. (1987). DNA sequences homologous to the drosophilia *opa* repeat are present in murine mRNAs that are differentially expressed in fetuse and adult tissues. *Mol. Cell. Biol.* **7**: 2003–2006.

Dyken, P. R., and Harper, P. S. (1973). Congenital Myotonica Dystrophica. *Neurology* **23**: 465–473.

Edwards, A., Civitello, A., Hammond, H. A., and Caskey, C. T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am. J. Hum. Genet.* **49**: 746–756.

Edwards, A., Hammond, H. A., Caskey, T. C., and Chakraborty, R. (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* **12**: 241–253.

Epplen, J. T. (1988). On simple repeat GATA/GACA sequences: A critical reappraisal. *J. Hered.* **79**: 409–417.

Fu, Y-H., Kuhl, D. P., Pizzuti, A., Pieretti, M., Sutcliffe, J. S., Richards, S., Verkerk, A. J., Holden, J. J., Fenwick, R. G., Jr., and Warren, S. T. (1991). Variation of the CGG repeat at the fragile X site results in genetic instability: Resolution of the Sherman paradox. *Cell* **67**: 1047–1058.

Gonzalez, I. L., Gorski, J. L., Campen, T. J., Dorney, D. J., Erickson, J. M., Sylvester, J. E., and Schmickel, R. D. (1985). Human 28s ribosomal RNA gene. *Proc. Natl. Acad. Sci. USA* **82**: 7666–7670.

Gottesman, I. I. (1991). "Schizophrenia Genesis. The Origins of Madness," pp. 96, Freeman, New York.

Guerini, D., and Klee, C. B. (1989). Cloning of human calcineurin A: Evidence for two isozymes and identification of a polyproline structural domain. *Proc. Natl. Acad. Sci. USA* **86**: 9183–9187.

Harley, H. G., Brook, J. D., Rundle, S. A., Crow, S., Reardon, W., Buckler, A. J., Harper, P. S., Housman, D. E., and Shaw, D. J. (1992a). Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature* **355**: 545–546.

Harley, H. G., Rundle S. A., Reardon, W., Myring, J., Crow, S., Brook, J. D., Harper, P. S., and Shaw, D. J. (1992b). Unstable DNA sequence in myotonic dystrophy. *Lancet* **339**: 1125–1128.

Haynes, S. R., Rebbert, M. L., Mozer, B. A., Forquignon, F., and David, I. B. (1987). *pen* repeat sequences are GGN clusters and encode a glycine-rich domain in a Drosophila cDNA homologous to the rat helix destabilizing protein. *Proc. Natl. Acad. Sci. USA* **84**: 1819–1823.

Howeler, C. J., Busch, H. F. M., Geraedts, J. P., Niermeijer, M. F., and Staal, A. (1989). Anticipation in myotonic dystrophy: Fact or fiction? *Brain* **112**: 779–797.

Ivanova, M. N., Frolova, E. I., and Georgiev, G. L. (1984). "Simple" sequences of the rat genome detected by hybridization with adenovirus DNA. *Doklady Akadem. Nauk. SSR* **276**: 984–988.

Kadonaga, J. T., Courey, A. J., Ladika, J., and Tjian, R. (1988). Dis-

tinct regions of Sp1 modulate DNA binding and transcriptional activation. *Science* **242**: 1566–1570.

Kao, C. C., Liebrman, P. M., Schmidt, M. C., Zhou, Q., Pei, R., and Berk, A. J. (1990). Cloning of a transcriptionally active TATA binding factor. *Science* **248**: 1648–1650.

Kendler, K. S., Gruenberg, A. M., and Tsuang, M. T. (1985). Psychiatric illness in first degree relatives of schizophrenic and surgical control patients: A family study using DSM III criteria. *Arch. Gen. Psychiat.* **42**: 770–779.

Kremer, E. J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S. T., Schlessinger, D., Sutherland, G. R., and Richards, R. I. (1991). Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)*n*. *Science* **252**: 1711–1714.

Lane, L. K., Shull, M. M., Whitmer, K. R., and Lingrel, J. B. (1989). Characterization of two genes for the human Na, K-ATPase beta subunit. *Genomics* **5**: 445–453.

La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E., and Fischbeck, K. H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* **352**: 77–79.

Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Jansen, G., Neville, C., Narang, M., Barcelo, J., and O'Hoy, K.(1992). Myotonic dystrophy mutation: An unstable CTG repeat in the 3′ untranslated region of the gene. *Science* **255**: 1253–1255.

McInnis, M. G., McMahon, F. J., Chase, G. A., Simpson, S. G., Ross, C. A., and DePaulo, J. R. (1993). Anticipation in bipolar affective disorder. *Am. J. Hum. Genet.*, in press.

Mermod, N., Williams, T., and Tjian, R. (1988). Enhancer binding factors AP-4 and AP-1 act in concert to activate SV40 late transcription *in vitro*. *Nature* **332**: 557–561.

Ohno, S., Emori, Y., and Suzuki, K. (1986). Human mRNA for calcium dependent protease (small unit). *Nucleic Acids Res.* **14**: 5559.

Richards, R. I., and Sutherland, G. R. (1992). Dynamic mutation: A new class of mutations causing human disease. *Cell* **70**: 709–712.

Ridley, R. M., Frith, C. D., Crow, T. J., and Conneally, P. M. (1988). Anticipation in Huntington's disease is inherited through the male line but may originate in the female. *J. Med. Genet.* **25**: 589–595.

Riggins, G. J., Lokey, L. K., Chastain, J. L., Leiner, H. A., Sherman S. L., Wilkinson, K. D., and Warren, S. T. (1992). Human genes containing polymorphic trinucleotide repeats. *Nature Genet.* **2**: 186–191.

Ross, C. A., McInnis, M. G., Margolis, R. C., and Li, S. H. (1993). Genes with triplet repeats: Candidate mediators of neuropsychiatric disorders. *Trends Neurosci.*, in press.

Sambrook, J., Fritsch, E. F., and Mauniatis, T. (1989) "Molecular Cloning: A Laboratory Manual," 2nd Ed., Cold Spring Harbor Laboratory Press, New York.

Sherman, S. L., Jacobs, P. A., Morton, N. E. Froster-Iskenius, U., Howard-Peebles, P. N., Nielsen, K. B., Partington, N. W., Sutherland, G. R., Turner, G., and Watson, M. (1985). Further segregation analysis of the fragile X syndrome with special reference to transmitting males. *Hum. Genet.* **69**: 3289–3299.

Tautz, D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* **17**: 6463–6471.

Tautz, D., and Renz, M. (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids. Res* **12**: 4127–4138.

Theune, S., Fund, J., Todd, S., Sakaguchi, A. Y., and Naylor, S. L. (1991). PCR primers for human chromosomes: Reagents for the rapid analysis of somatic cell hybrids. *Genomics* **9**: 511–516.

Tilley, W. D., Marcelli, M., Wilson, J. D., and McPhaul, M. J. (1989). Characterization and expression of a cDNA encoding the human androgen receptor. *Proc. Natl. Acad. Sci. USA* **86**: 327–331.

Tsilfidis, C., MacKenzie, A. E., Mettler, G., Barcelo, J., and Korneluk, R. G. (1992). Correlation between CTG trinucleotide repeat length and frequency of severe congenital myotonic dystrophy. *Nature Genet.* **1**: 192–195.

Verkerk, A. J., Pieretti, M., Sutcliffe. J. S., Fu, Y. H., Kuhl, D. P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M. F., and Zhang, F. P. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**: 905–914.

Warburton, D., Gersen, S., Yu, M. T., Jackson, C., Handelin, B., and Housman, D. (1990). Monochromosomal rodent–human hybrids from microcell fusion of human lymphoblastoid cells containing an inserted dominant selectable marker. *Genomics* **6**: 358–366.

Warren, A. C., McInnis, M. G., Blaschak, J., Kaliatsidaki, M., Petersen, M. B., Chakravarti, A., and Antonarakis, S. E. (1992). Dinucleotide repeat (GT)n markers on chromosome 21. *Genomics* **14**: 818–819.

Weber, J. L. (1990). Informativeness of human (dC–dA)*n* (dG–dT) polymorphisms. *Genomics* **7**: 524–530.

Warburton, D., Gersen, S., Yu, M. T., Jackson, C., Handelin, B., and Housman, D. (1990). Monochromal rodent–human hybrids from microcell fusion of human lymphoblastoid cells containing an inserted dominant selectable marker. *Genomics* **6**: 358–366.

Wharton, K. A., Yedvobnick, B., Finnerty, V. G., and Artvanis-Tsakonas, S. (1985). *opa:* A novel family of transcribed repeats shared by the Notch locus and other developmentally regulated loci in D. melanogaster. *Cell* **40**: 55–62.

Wong, Z., Wilson, V., Patel, I., Povey, S., and Jeffreys, A. J. (1987). Characterization of a panel of highly variable minisatallite clones from human DNA. *Ann. Hum. Genet.* **51**: 269–288.

Yu, S., Mulley, J., Loesch, D., Turner, G., Donnelly, A., Gedeon, A., Hillen, D., Kremer, E., Lynch, M., and Pritchard, M. (1992). Fragile-X syndrome: Unique genetics of the heritable unstable element. *Am. J. Hum. Genet.* **50**: 968–980.

Zischler, H., Kammerbauer, C., Studer, R., Grzeschik, K. H., and Epplen, J. T. (1992). Dissecting $(CAC)_5/(GTG)_5$ multilocus fingerprints from man into individual locus-specific, hypervariable components. *Genomics* **13**:983–990.

Exhibit D

Genes with triplet repeats: candidate mediators of neuropsychiatric disorders

Christopher A. Ross, Melvin G. McInnis, Russell L. Margolis and Shi-Hua Li

Christopher A. Ross, Melvin G. McInnis, Russel L. Margolis and Shi-Hua Li are at the Laboratory of Molecular Neurobiology, Depts of Psychiatry and Neuroscience, Johns Hopkins University School of Medicine, 720 Rutland Ave, The Ross Building, Room 618, Baltimore, MD 21205-2196, USA.

Recently a new form of human mutation – expansion of trinucleotide repeats – has been found to cause the diseases of fragile X syndrome, spinal and bulbar muscular atrophy, myotonic dystrophy and, most recently, Huntington's disease. We review the emerging data on the genetics and neurobiology of these disorders. Three are characterized by unusual patterns of inheritance, in particular, genetic 'anticipation', in which the severity of the disorder increases and the age of onset decreases in successive generations of a pedigree. Several idiopathic neuropsychiatric disorders have features of inheritance consistent with anticipation. In bipolar affective disorder, there is evidence for both earlier age of onset and more severe illness in the second generation of a subset of unilineal pedigrees. There is also the suggestion of anticipation in some forms of schizophrenia, spinocerebellar atrophy and autism. Triplet repeats are present in additional known genes, both in coding regions and untranslated regions. Furthermore, many novel genes with triplet repeats are expressed in the human brain, and these are candidates to cause some forms of these neuropsychiatric disorders.

Trinucleotide (or triplet) repeats are repeated units of three nucleotides present in genomic DNA. When they occur in the coding regions of genes, they can direct the synthesis of repeating single amino acids. The most common triplet repeats in human DNA included in the Genbank database are CGG, CAG, CAA, TAA and GAG (Ref. 1). (Note that in double-stranded DNA, any triplet could be read in six possible reading frames; CAG is equivalent to AGC, GCA, CTG, TGC or GCT.) Since the lengths of these repeats (like those of other microsatellites, or short repeating sequences of DNA) at any given locus are often polymorphic – variable from individual to individual – they are useful as linkage markers in the Human Genome Project. Otherwise, they appeared to be curiosities, until four neuropsychiatric disorders were recently discovered to be caused by expansion of unstable trinucleotide repeats (recently reviewed by Davies[2], Richards and Sutherland[3] and Caskey *et al.*[4]).

Myotonic dystrophy

Myotonic dystrophy is an autosomal dominant disorder characterized primarily by myotonia and progressive muscle weakness, and is the most common form of adult-onset muscular dystrophy[5]. There may be cognitive changes, including mental retardation, as well as skeletal, cardiovascular and ocular manifestations. The disorder shows 'anticipation', with severity increasing over three or four generations, presenting as cataracts or mild alteration in muscle contractility in late adulthood of the first generation, moderate weakness in an intermediate generation, and progressing to a more severe disease state including neonatal onset with mental retardation in a later generation[6]. These aspects of inheritance were so unusual that they were attributed to artifacts of ascertainment bias (probands coming to medical attention tend to present a 'late-generation' form of the disorder and so are more severely affected than other members of their family) or other artifacts[7], and only recently have been clinically confirmed as anticipation[8].

The gene that causes myotonic dystrophy, called DM-1 or myotonin protein kinase, has recently been identified using positional cloning and, in one case, direct screening for triplet repeats; this gene contains a CTG repeat in its 3′ untranslated region (see Fig. 1)[9–11]. The length of the CTG repeat sequence is highly polymorphic in the normal population, with a range of 5 to 30 repeats. Patients have expanded repeats, often with many hundreds of repeats, and the extent of expansion correlates with severity of the disease[12,13]. The gene has regions of strong homology to cAMP-dependent protein kinase[10,11] and is expressed in brain, heart and muscle[10].

Fragile X syndrome

Fragile X syndrome is the second most common genetic form of mental retardation (after Down syndrome), occurring in about one of 2000 births. It was originally identified by the association of a folate-sensitive fragile site on the distal end of the long arm of the X chromosome (band Xq27.3) with mental retardation. Hemizygous males have the fragile X locus on their single X chromosome and generally have moderate to severe retardation, often with pervasive developmental delays and autistic behavior[4,14]. Physical signs may include alterations such as testicular enlargement, anatomical defects and connective tissue abnormalities.

The pattern of inheritance in families with the fragile X marker shows a form of anticipation, with increasing penetrance in succeeding generations. Passage through a female increases the risk in the next generation for children to be affected. These two characteristics comprise the so-called 'Sherman paradox' – the first-degree male relatives of transmitting males have less chance of being affected than male grandchildren of such males[5,14,15]. Obligate female carriers (some of whom presumably have the full mutation) of the fragile X chromosome are not unaffected, but may only have mild cognitive defects and psychiatric features such as schizotypal symptoms[16].

The genetic basis of the fragile X syndrome is the expansion of a triplet repeat in a gene called FMR-1 (see Fig. 1)[4,17–21]. The full-length sequence for the transcribed portion of the gene has not yet been published (it is a 4.8 kb mRNA). Within what is believed to be the 5′ untranslated portion of the message is a region of multiple repeats of CGG. In the normal population, the length of the repeat is highly polymorphic, ranging between 6 and about 42 triplets. In pedigrees with the fragile X syndrome, the CGG repeat can become unstable and then its

Ex. D

length expands during transmission from one generation to the next. Unaffected members of the family, like members of the general population, have less than 50 repeats, or may carry a so-called 'premutation' with 50–200 repeats. The premutation is unstable, and can expand to an extent sufficient to cause disease in subsequent generations. Affected members may have expansions of up to several thousand bases. The size of the premutation in women correlates with their likelihood of bearing affected children.

In the normal population, mRNA of the FMR-1 gene is expressed at the highest levels in the brain and the testes, two of the major organs affected in the fragile X syndrome. Expression in the brain appears primarily neuronal and generally parallels neuronal density, with high levels found in the cerebellar granule cell layer, hippocampus and some or all neurons in cerebral cortex[22]. In addition, relatively high levels of expression are present in ovarian follicle and the eye, and the mRNA is also widely expressed in early embryonic development. The function of the FMR-1 gene is unknown. It has no homology to other known sequences, except in the area of the repeats itself. The predicted protein structure does contain a region coding for a sequence of amino acids (KKXK) that might cause nuclear translocation of the protein, suggesting the gene could code for a DNA-binding protein or some other nuclear regulatory protein.

Spinal and bulbar muscular atrophy

X-Linked spinal and bulbar muscular atrophy (also called Kennedy's disease) is a rare, late-onset form of motor neuron degeneration[23] that may be accompanied by mental retardation and partial insensitivity to androgens. Because the human androgen receptor gene is located on the X chromosome, it seemed possible that it represented the site of mutation and so was studied in patients with this disorder. An expanded CAG repeat was found; patients had numbers of repeats ranging from 40 to 52 while controls had a range of only 17 to 26 (Ref. 24). There was no overlap in the number of triplet repeats between the patients and control samples, and no alteration in any other region of the sequence of the androgen receptor was found, strongly suggesting that expansion of the triplet repeat accounts for the disease.

Like other steroid receptors, the androgen receptor is a DNA-binding protein involved in regulation of transcription. The CAG repeat is in the first exon, outside of the region that encodes either DNA- or hormone-binding activity[24]. In the reading frame of the androgen receptor, CAG repeats are presumably translated into a polyglutamine tract. The phenomenon of anticipation has not been clearly observed in spinal and bulbar muscular atrophy, but the degree of expansion does correlate with severity of the illness[25,26].

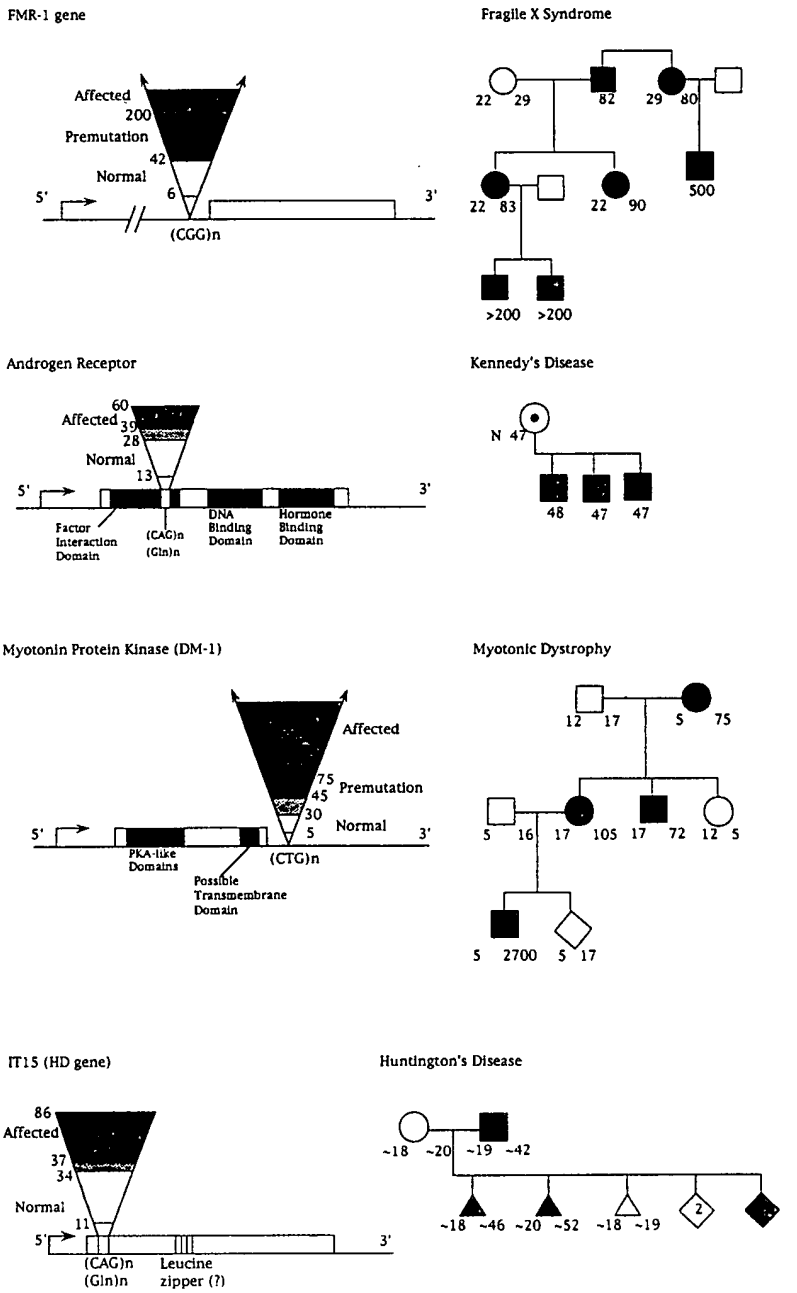


Fig. 1. *Diseases caused by expansion of trinucleotide repeats. On the left, a schematic diagram of each gene, with the repeat region depicted as an inverted pyramid and the numbers of repeats beside it. Open regions represent the normal variation of repeat number; blue regions represent the range of abnormally long repeat sequences that are unstable but not causative of prominent clinical symptoms (premutations); and red regions represent expanded repeat lengths that cause disease. A representative pedigree for each disorder is depicted to the right. Open, blue and red symbols represent unaffected individuals, individuals with premutations, and affected individuals, respectively. The number of repeats in each allele is displayed below the symbol. Circles represent females, squares represent males, diamonds or triangles represent individuals whose sex is not identified in order to preserve confidentiality, with a number inside to indicate more than one person. A dot inside a circle indicates a carrier.* (Gene diagrams are adapted from Ref. 3, and pedigrees are from Refs 4, 25 and 31.)

Huntington's disease

Huntington's disease is an autosomal dominant syndrome linked to chromosome 4p, involving movement disorder, emotional disorder and dementia[27]. Affected children of affected fathers have an age of onset eight to ten years earlier than

their fathers, while affected children of affected mothers have an onset age similar to their mothers. This phenomenon appears to continue through the generations, with some children of affected fathers having juvenile onset[28,29]. These features led us to suggest[30] that Huntington's disease might be caused by expansion of a triplet repeat in a gene on chromosome 4.

Very recently, the Huntington's disease Collaborative Research Group has used exon trapping (a technique for extracting exons from genomic DNA cloned in vectors such as cosmids) to identify a gene from chromosome 4p, termed IT15 ('interesting transcript 15'), containing a repeating CAG triplet[31,87]. Normals have between 11 and 34 repeats with a median of 19, while patients with Huntington's disease have between 37 and 86 copies with a median of 45. The CAG repeats appear to be in the 5′ portion of the coding region, which yields a protein with a deduced molecular weight of about 348 kDa. In the predicted reading frame, the CAGs code for polyglutamine, like in the androgen receptor gene. The gene appears to be widely expressed in brain and peripheral tissues as an 11 kb transcript. Patients with longer repeats have an earlier age of onset with a high correlation ($r = 0.75$) between length of repeating region and age of onset[87,88]. However, for most Huntington's disease alleles (which have between 37 and 52 repeats) there is such a wide range of ages of onset that the length of the repeat cannot be used to predict the age of onset for individual patients. The length of the Huntington's disease allele is highly unstable in vertical transmission, with a difference between the two sexes. During female transmission, the length of the Huntington's disease allele frequently changes, but only by a few repeats, so that the mean length of the allele in the child is very similar to that in the mother. However, in paternal transmission, the repeat lengthens dramatically in about one third of the cases to about twice that of the paternal allele. This appears to occur during spermatogenesis[87].

Pathophysiology

These diseases represent a new class of disorders caused by alterations in unstable DNA sequences. Thus, rather than having fixed mutations these diseases are caused by so-called 'dynamic' mutations[3] that can change in each generation, accounting for the unusual aspects of inheritance, especially anticipation. The mechanism by which the triplet repeats expand is unknown. Unequal crossing over during recombination could not account for the large expansions from hundreds to thousands of bases that are observed. There is an enzyme (telomerase) that inserts multiple copies of a six-base-repeat into telomeric DNA (Ref. 32), but no enzymes that could insert a three-base-repeat into genomic DNA are currently known. It has been proposed that slippage of a DNA polymerase during DNA replication may be involved[33].

In myotonic dystrophy and fragile X syndrome there appear to be two separate mutational events. In the normal population, the length of the triplet repeat is polymorphic, but stably inherited. The first event is the formation of a 'premutation' or allele with a slightly longer repeat, which appears to be a rare event since founder chromosomes (haplotype evidence of a single event on an ancestral chromosome) have been suggested for both disorders[34,35]. Population genetic modeling of the fragile X syndrome suggests that the premutation persists in the population for a mean of 60 generations[36,37]. The premutation renders the DNA sequence unstable such that it can then expand to a much greater length, and cause disease. Once the expansion has taken place, it becomes increasingly unstable – the longer the expansion, the more it can expand in subsequent generations, with the cycle presumably ending when the disease becomes so severe as to prevent reproduction. Rare contractions of an expanded allele back to normal length (along with normal phenotype) have been described for myotonic dystrophy[38,39]; the mechanism may involve a gene conversion event[40].

In both myotonic dystrophy and fragile X syndrome, the repeat is more likely to expand during a female meiosis. In fragile X syndrome there appears to be a relationship between methylation of a CpG island in the region 5′ to the gene and expansion of the repeat. Which comes first is currently uncertain[41], though indirect evidence suggests that the expansion is the primary event[21,42]. In Kennedy's disease, by contrast, the expansion behaves more like a premutation seen in the other disorders; large expansions do not occur, but there is mild instability of the length of the repeats, which can yield both small expansions and small contractions[25,43]. In this disease the repeat length appears more unstable in male meiosis.

It is uncertain how the expanded triplet repeat causes these diseases. Large expansions may make it impossible for the message to be transcribed or translated, possibly related to the hypermethylation of the CpG island. Pieretti *et al.*[44] have found that the FMR-1 mRNA is absent in the majority of male fragile X patients. In the case of fragile X syndrome, simple absence of the message may cause symptoms, since other mutations in the same gene, such as deletions[45] or a point mutation[46], appear to cause a phenotype similar to fragile X without expansion of the CGG repeat, presence of the folate-sensitive fragile site or methylation of the CpG island.

To our knowledge, no pathological studies of the brains of fragile X patients have yet been reported. A study using magnetic resonance imaging (MRI) found selective atrophy of the cerebellar vermis in fragile X patients compared to either normals or patients with other developmental disabilities[47]. How this relatively selective effect results from a mutation in a gene expressed widely in the brain[22] is unclear.

In the case of myotonic dystrophy, mRNA levels for the expanded allele are decreased, dramatically in congenital cases, while levels of the mRNA corresponding to the normal allele appear unchanged[48,49]. Overall protein levels are modestly

decreased. How these changes cause the dominance of the gene is unclear, but it may be a dosage effect[49]. Normal expression of the DM-1 gene is just beginning to be defined (Schalling, G., pers. commun.). Like the FMR-1 message, it is widely expressed in brain, but apparently with more regional selectivity; for instance, it appears to be relatively highly expressed in cerebellar Purkinje cells. Several splice isoforms of the DM-1 mRNA have been described, but their biological significance is as yet unknown[49,50]. The means by which expansion of repeats in the 3′ untranslated region yields decreased levels of message is not clear, but this region in at least some genes may have an influence on regulation of mRNA (Ref. 51).

In the case of the gene for the androgen receptor, which has the repeat in a coding region, the effect of the expanded triplet repeat on protein function has been studied[52], in cell culture, with normal or expanded androgen receptor cDNA and a reporter gene coupled to the androgen receptor response element. Although androgen binding was normal, there was decreased transactivation of the hormone response element. Deletion of the region with the polyglutamine tract does not affect receptor function as much as the expansion. Mutations in other parts of the molecule, especially the ligand-binding region, by contrast, cause a different disorder – the androgen insensitivity syndrome[53,54]. The N-terminal region of the androgen receptor has recently been implicated in binding to other cell-specific enhancer factors[55]. Expansion of the triplet repeat could prevent the androgen receptor from interacting with some other enhancer factor that has a restricted cellular expression pattern, which might, in part, explain the selective neuronal vulnerability observed in the disorder.

As yet, little is known about the mechanism by which the expanded triplet repeat in IT15 causes Huntington's disease, a fully penetrant dominant disorder. It may be that the greater prevalence of anticipation in male transmission relates to some form of imprinting. It is uncertain how a mutation in a gene coding for a protein expressed widely in the brain and body causes neuronal degeneration preferentially in the basal ganglia and cerebral cortex. Preliminary data[88] suggest that the IT15 message is expressed in neurons and to a lesser extent in glia throughout the brain. Expression in the brain is higher than that in the periphery, but does occur widely in many tissues in the body. Furthermore, expression does not appear to be altered dramatically in either the caudate nucleus or cortex of Huntington's disease patients compared with controls. Thus the pathophysiology of the disorder might involve an action of an abnormal IT15 protein product. The protein encoded by IT15 has no prominent hydrophobic domains, and thus may be soluble. It might be a structural protein, or a protein that interacts in a non-covalent manner with a structural protein. It has one leucine zipper (of four leucines) and so could be a transcription factor of the Fos/Jun family, though it would be much larger than any other members of the family[56]. If it were a transcription factor, it is conceivable that, in a manner similar to the androgen receptor, expansion of the polyglutamine repeat could alter its interaction with a second, cell-type-specific factor, causing a gain-of-function phenotype with a restricted anatomical distribution corresponding to that of the second factor. Similarly, if it were a structural protein, an interaction with other structural proteins might be disrupted. An alternative hypothesis suggests that Huntington's disease might be related to a mitochondrial dysfunction, perhaps inducing excitotoxicity[57].

TABLE I. Genes containing CAG repeats[a]

Gene name	Perfect trinucleotide repeats	Amino acid repeats
Androgen receptor* (M21748)	CAG (× 17)	Gln (× 21), glutamine-rich region
IT15 (Huntington's gene)* (L12392)	CAG (× 21)	Gln (× 23), glutamine-rich region
TATA-binding protein (M55654/M34960)	CAG (× 18)	Gln (× 38), glutamine-rich region
Transcription factor TFEB (M33782)	CAG (× 6)	Gln (× 10), glutamine-rich region
Transforming growth factor β (M60315)	CAG (× 6)	Gln (× 6)
Protective protein (M22960)	CTG (× 8)	Leu (× 9), signal sequence
Placental alkaline phosphatase (M12551)	CTG (× 7)	Leu (× 7), signal sequence
Zinc-finger protein (ETR 103) (M80583/M62829)	AGC (× 6)	Ser (× 6), serine- and glycine-rich region
Natriuretic peptide (M31776)	CAG (× 7)	5′ untranslated region
pim-1 Proto-oncogene (M27903)	CAG (× 8)	5′ untranslated region
Erythrocyte 2,3-bisphosphoglycerate mutase (M04327)	CTG (× 8)	5′ untranslated region
Myotonin kinase* (M87313)	CAG (× 5) (13)[b]	3′ untranslated region
Proenkephalin B (X00177)	CAG (× 7)	3′ untranslated region

[a] Genbank (release 71.0) was searched with a 30-base string of ten CTG repeats (CAG equivalent to AGC, GCA, CTG, TGC or GCT). All genes shown have no more than two mismatches and a perfect match of at least six repeats. (The number of matches is far greater than would be expected by chance: a search with an arbitrary 30-base-long string yields no better than 13 mismatches.)
[b] The sequence in the database at the time the search was done had five repeats, but the most common allele has 13 repeats.
* Genes which are known to undergo triplet repeat expansion, causing disease.

Other disorders with anticipation

The discovery in rapid succession of four diseases caused by expansion of triplet repeats raises the possibility that more remain to be found. Several additional neuropsychiatric disorders with clinical features of anticipation could be candidates.

Perhaps the most striking possibility is bipolar affective illness, which is a much more common disease than any of the previously discussed

TABLE II. Genes containing GAG repeats[a]

Gene name	Perfect trinucleotide repeats	Amino acid repeats
80K-H protein (protein kinase C substrate) (J03075)	GAG (× 9)	Glu (× 12)
Histidine-rich calcium-binding protein (M60052)	GAG (× 9)	Glu (× 12)
Cyclin D (M64349)	GAG (× 8)	Glu (× 9)
Chromogranin A (J03483)	GAG (× 8)	Glu (× 9)
hCENP-B gene for centromere autoantigen B (X55039)	GAG (× 6)	Glu (× 6), near 14 consecutive glu
α_2 Adrenergic receptor (M34041)	GAG (× 6)	Glu (× 12)
Arginine-rich gene (M8375)	AGG (× 7)	Arg (× 8)
Acid (type I) cytokeratin 10 (KRT10) (X14487)	GGA (× 8)	Gly (× 8)
γ-Glutamylcysteine synthetase (M90656)	GAG (× 9)	5′ untranslated region
trk Oncogene (X03541)	(GAG) (× 6)	5′ untranslated region, GA-rich region

[a]Genbank search was performed in the same manner as for CAG repeats. GAG is equivalent, for purposes of the Genbank search, to AGG, GGA, CTC, TCC and CCT.

disorders, with a lifetime prevalence of about 0.5–1% of the population. In bipolar affective illness, there appears to be genetic transmission with an as yet undetermined pattern of inheritance in many families[58]. We have studied pedigrees with unilineal transmission of bipolar affective illness[59], finding marked anticipation, with age of onset about 11 years earlier and episode frequency about twice as great in the second generation compared with the first[60]. A Cox proportionate hazard analysis showed that the anticipation was separable from the previously known changing rate of major depression during this century or 'cohort effect'[61]. In addition, changing patterns of drug abuse, deaths of affected persons prior to interview, decreased fertility and censoring of age of onset did not account for the results. Anticipation appears to be present in about 40% of the pedigrees.

Schizophrenia also runs in families, with an uncertain pattern of inheritance. The presence of individuals with 'schizophrenia spectrum' disorders in families of probands with schizophrenia[62] suggests variable penetrance. The risk to parents of probands is 6%, to siblings 9% and to children of probands 12%, combining data taken from a number of European studies of the illness using consistent diagnostic criteria[63]. This pattern could have other interpretations, but is consistent with anticipation. Review of age of onset data from a study of pedigrees[64] with more than one affected case is also suggestive of anticipation. In about 15 families, there was anticipation of 10 years or greater in transmitting pairs (Pulver, A., pers. commun.).

Autosomal dominant hereditary ataxia also displays variable age of onset, with some evidence for anticipation. This heterogeneous group of disorders, also referred to as spinocerebellar ataxia or hereditary olivo-ponto-cerebellar atrophy, is characterized by cerebellar ataxia, variably accompanied by dementia and extrapyramidal signs. In 1950, Schut[65] explicitly pointed out the presence of anticipation in a large pedigree, noting both a younger age of onset and a more severe course in later generations. Although no study specially designed to exclude ascertainment bias has been performed, data from a number of other pedigrees also suggest the presence of anticipation in at least some forms of hereditary ataxia[66–68].

Finally, the distribution of illness in families of patients with autism not associated with fragile X syndrome could also be consistent with anticipation with a dominant pattern of inheritance. About half of parents of autistic probands have mild neuropsychiatric disabilities that are characterized by social language disorders and aloofness[69].

Functions of genes with triplet repeats

In addition to the four genes causing diseases described above, other genes with triplet repeats are already known. There are about a dozen human genes with CGG repeats already entered into Genbank[3,70]. In several cases the length of the repeat is polymorphic in the normal population. In most cases, CGG repeats appear to be in the 5′ untranslated regions of the transcripts[3]. Unlike the CGG repeats, CAG repeats are predominantly within coding regions, and glutamine is the most common amino acid encoded (Table I). For GAG repeats in genes, eight out of ten are in coding regions, with glutamate being the most common amino acid (Table II). Another common triplet repeat, TAA, has been reported in only two genes, both in untranslated regions (not shown).

The functions, if any, of repeated single amino acids in a protein are uncertain. Different amino acid repeats may have different functions. Repeated amino acids, especially glutamine, are present in a number of transcription factors, most likely in non-DNA-binding regions. Proteins containing so-called Opa repeats, which are polyglutamine tracts, in *Drosophila* and mouse[71,72] often code for developmental-stage-specific transcription factors including homeodomain proteins. Human TATA-binding protein has 38 consecutive glutamine residues in its N-terminal domain[73]. Transcription factor Sp1 has a glutamine-rich region with shorter glutamine repeats, which are important for transcription[74,75]. Repeat sequences of GGN, known as Pen, isolated in *Drosophila*, encode polyglycine and are present in a homologue of the rat helix-destabilizing protein[76]. A proline-rich transcriptional activation zone has also been described[77]. Regions rich in acidic amino acids are activational domains in a number of transcription factors[78]. Inducible transcription factors would be interesting candidate genes for brain disorders, because of their involvement in development and plasticity[79–81].

Other repeats may be involved in different protein–protein interactions. For instance, a repeat of 11 proline residues in human calcineurin A has been proposed to mediate binding to calmodulin[82]. In some cases, repeats of the hydrophobic amino acid leucine are present in signal sequences, where the exact amino acid sequence is believed to be less important than the overall hydrophobicity.

Additional novel genes with triplet repeats

Knowing that triplet-containing genes are present in the human genome[83], and positing the possibility that other diseases could be caused by expansion of triplet repeats, the question arises: are there additional genes with triplet repeats yet to be discovered? A number of groups have begun the search for such genes[30,70,84–86]. One approach involves direct screening of patients' genomic DNA for expanded repeats using oligonucleotides and a ligase cycle reaction, termed 'repeat expansion detection'[84]. This method can detect the expansions in myotonic dystrophy and fragile X DNA, and has suggested the presence of an expanded CTG repeat on chromosome 18 in three families of the Centre d'Etude du Polymorphisme Humaine (CEPH).

Another approach is to screen cDNA libraries to identify individual candidate clones. Riggins *et al.*[70] recently identified a number of genes with triplet repeats by this method. Because they screened at relatively low stringency, the repeats identified were short (all but one had five or fewer repeats) and therefore were only slightly polymorphic in the normal human population. Since the four known diseases caused by expansion of triplet repeats all have repeat lengths that are highly polymorphic in the normal population, one would like a screening process that can identify repeats long enough to have such polymorphisms.

By screening human brain cDNA libraries at high stringency we have been able to identify a series of novel genes with triplet repeats[30]. Fourteen out of an initial group of forty clones encoded novel human genes, and eight of the inserts have been fully sequenced. All contain repeats and five of the eight have nine or more perfect repeats (see Fig. 2). The first clone sequenced was the myotonic dystrophy gene with thirteen CTG repeats. In five clones with CAG repeats, the repeats were within long open reading frames, suggesting that they may be translated into amino acids. In four of these cases the amino acids would be glutamine.

Northern blot analyses show that all of the genes are expressed in brain, with the suggestion of differential distribution (some were expressed more highly in cerebral cortex than cerebellar cortex, while others showed the reverse). Chromosomal assignments reveal a distribution on several autosomes and one on the X chromosome. Furthermore, the repeat lengths of four out of the six CAG triplets were polymorphic, with two very highly polymorphic, one with 85% heterozygosity and thirteen different alleles.

In conclusion, expansion of triplet repeats constitutes a new form of dynamic mutation capable of causing diseases with unusual forms of inheritance, particularly anticipation. A number of neuropsychiatric disorders, especially bipolar affective disorder, show features suggestive of anticipation. Furthermore, there appear to be many genes with triplet repeats expressed in the human brain. Some of the repeats are relatively long and polymorphic in the normal population. We therefore suggest that novel human genes with triplet repeats may be candidates for these neuropsychiatric disorders.

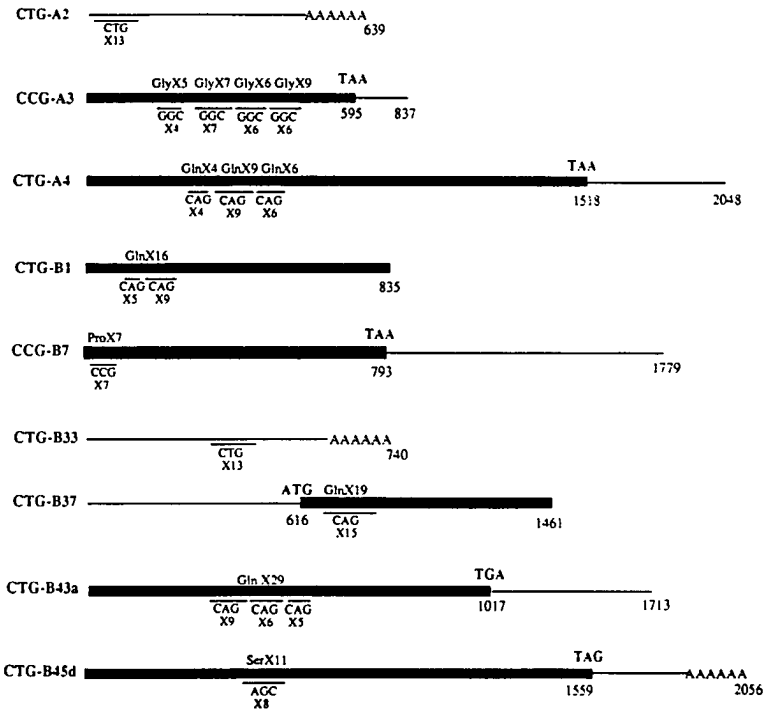


Fig. 2. *Maps of the inserts of nine novel clones containing triplet repeats, obtained by screening a human brain cDNA library with CTG or CCG repeat oligonucleotides[30]. Open reading frames are indicated by thicker lines. Nucleotide repeats are shown below, and predicted amino acids above the lines. Note that most of the repeats appear to be in open reading frames. The first clone sequenced (A2) was DM-1 of myotonic dystrophy, which was a novel sequence at the time.*

Note added in proof. A gene on chromosome 2 has recently been linked to familial colon cancer[89]. These cancers contain multiple instabilities in simple sequence repeats, including triplet repeats [identified using our probe CTG-B37 (Ref. 30 and Fig. 2)], which may contribute to the phenotype[90]. A genetic approach may uncover mechanisms underlying instability of triplet repeats.

Selected references

1 Beckmann, J. S. and Weber, J. L. (1992) *Genomics* 12, 627–631
2 Davies, K. E. (1992) *Nature* 356, 15
3 Richards, R. I. and Sutherland, G. R. (1992) *Cell* 70, 709–712
4 Caskey, C. T., Pizzuti, A., Fu, Y-H., Fenwick, R. G. and Nelson, D. L. (1992) *Science* 256, 784–788
5 Harper, P. S. (1989) *Myotonic Dystrophy* 2nd edn, W. B. Saunders

Acknowledgements

We thank Stylianos Antonarakis, Martin Schalling, Colin Stine, Be Wieringa, Ben Oostra, Leonard Pinsky, Colin Fletcher, Susan Chamberlain, Kenneth Fishbeck, James Gusella and Marcy MacDonald for helpful discussions or sharing data prior to publication, or both. We thank Solomon Snyder and Susan Folstein for reading a draft of the manuscript. Figure 1 is adapted with permission from figures by G. Sutherland, Thomas Caskey and Kenneth Fischbeck, and from data from the HD Collaborative Group. We thank Deborah Holifield for expert assistance with word processing.

6 Dyken, P. R. and Harper, P. S. (1973) *Neurology* 23, 465–473
7 Penrose, L. S. (1947) *Ann. Eugenet.* 14, 125–132
8 Höweler, C. J., Busch, H. F. M., Geraedts, J. P. M., Niermeijer, M. F. and Staal, A. (1989) *Brain* 112, 779–797
9 Mahadevan, M. *et al.* (1992) *Science* 255, 1253–1255
10 Brook, J. D. *et al.* (1992) *Cell* 68, 799–808
11 Fu, Y-H. *et al.* (1992) *Science* 255, 1256–1258
12 Harley, H. G. *et al.* (1992) *Nature* 355, 545–548
13 Tsilfidis, C., MacKenzie, A. E., Mettler, G., Barcelo', J. and Korneluk, R. G. (1992) *Nature Genet.* 1, 192–195
14 Hagerman, R. J. and Silverman, A. C. (1991) *Fragile X Syndrome*, Johns Hopkins University Press
15 Sherman, S. L. *et al.* (1985) *Hum. Genet.* 69, 3289–3299
16 Freund, L. S., Reiss, A. L., Hagerman, R. and Vinogradov, S. (1992) *Arch. Gen. Psychiat.* 49, 54–60
17 Fu, Y-H. *et al.* (1991) *Cell* 67, 1047–1058
18 Kremer, E. J. *et al.* (1991) *Science* 252, 1711–1714
19 Verkerk, A. J. M. H. *et al.* (1991) *Cell* 65, 905–914
20 Oberle, I. *et al.* (1991) *Science* 252, 1097–1102
21 Yu, S. *et al.* (1992) *Am. J. Hum. Genet.* 50, 968–980
22 Hinds, H. L. *et al.* (1993) *Nature Genet.* 3, 36–43
23 Kennedy, W. R., Alter, M. and Sung, J. H. (1968) *Neurology* 18, 671–680
24 La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E. and Fischbeck, K. H. (1991) *Nature* 352, 77–79
25 La Spada, A. R. *et al.* (1992) *Nature Genet.* 2, 301–304
26 Doyu, M. *et al.* (1992) *Ann. Neurol.* 32, 707–710
27 Folstein, S. E. (1989) *Huntington's Disease. A Disorder of Families*, Johns Hopkins University Press
28 Ridley, R. M., Frith, C. D., Crow, T. J. and Conneally, P. M. (1988) *J. Med. Genet.* 25, 589–595
29 Ridley, R. M., Frith, C. D., Farrer, L. A. and Conneally, P. M. (1991) *Med. Genet.* 28, 224–231
30 Li, S-H., McInnis, M. G., Margolis, R. L., Antonarakis, S. E. and Ross, C. A. (1993) *Genomics* 16, 572–579
31 Huntington's Disease Collaborative Research Group (1993) *Cell* 72, 971–983
32 Grieder, C. and Blackburn, R. (1987) *Cell* 51, 887–898
33 Schlötterer, C. and Tautz, D. (1992) *Nucl. Acids Res.* 20, 211–215
34 Richards, R. I. *et al.* (1992) *Nature Genet.* 1, 257–260
35 Harley, H. G. *et al.* (1991) *Am. J. Hum. Genet.* 49, 68–75
36 Morton, N. E. and Macpherson, J. N. (1992) *Proc. Natl Acad. Sci. USA* 89, 4215–4217
37 Chakravarti, A. (1992) *Nature Genet.* 1, 237–238
38 Shelbourne, P. *et al.* (1992) *Hum. Mol. Genet.* 1, 467–473
39 Brunner, H. G. *et al.* (1993) *New Engl. J. Med.* 328, 476–480
40 O'Hoy, K. L. *et al.* (1993) *Science* 259, 809–812
41 Hansen, R. S., Gartler, S. M., Scott, C. R., Chen, S. H. and Laird, C. D. (1992) *Hum. Mol. Genet.* 1, 571–578
42 Bell, M. V. *et al.* (1991) *Cell* 64, 861–866
43 Biancalana, V. *et al.* (1992) *Hum. Mol. Genet.* 1, 255–258
44 Pieretti, M. *et al.* (1991) *Cell* 66, 817–822
45 Gedeon, A. K. *et al.* (1992) *Nature Genet.* 1, 341–344
46 DeBoulle, K. *et al.* (1993) *Nature Genet.* 3, 31–35
47 Reiss, A. L., Aylward, E., Freund, S. F., Joshi, K. and Byran, R. N. (1991) *Ann. Neurol.* 29, 26–32
48 Novelli, G. *et al.* (1992) *Am. J. Hum. Genet.* 51, A40
49 Fu, Y-H. *et al.* (1993) *Science* 260, 235–238
50 Jansen, G. *et al.* (1992) *Nature Genet.* 1, 261–266
51 Ahringer, J. and Kimble, J. (1991) *Nature* 349, 346–348
52 Pinsky, L. *et al.* (1992) *Am. J. Hum. Genet.* 51, Suppl. A41
53 Pinsky, L. *et al.* (1992) *Clin. Invest. Med.* 15, 456–472
54 Batch, J. A. *et al.* (1992) *Hum. Mol. Genet.* 1, 497–503
55 Adler, A. J., Danielsen, M. and Robins, D. M. (1992) *Proc. Natl Acad. Sci. USA* 89, 11660–11663
56 Landschulz, W. H., Johnson, P. F. and McKnight, S. L. (1988) *Science* 240, 1759–1764
57 Beal, M. F., Hyman, B. T. and Koroshetz, W. (1993) *Trends Neurosci.* 16, 125–131
58 Mendlewicz, J. and Rainer, J. D. (1974) *Am. J. Hum. Genet.* 26, 92–101
59 Simpson, S. G., Folstein, S. E., Meyers, D. A. and DePaulo, J. R. (1991) *Am. J. Psychiat.* 149, 1660–1665
60 McInnis, M. G. *et al. Am. J. Hum. Genet.* (in press)
61 Cross-National Collaborative Group (1992) *J. Am. Med. Assoc.* 268, 3098–3105
62 Kendler, K. S. and Gruenburg, A. M. (1984) *Arch. Gen. Psychiat.* 41, 555–564
63 Gottesman, I. I. (1991) in *Schizophrenia Genesis*, p. 96, Freeman
64 Pulver, A. E. *et al.* (1992) *Br. J. Psychiat.* 160, 65–71
65 Schut, J. W. (1950) *Arch. Neurol. Psychiat.* 63, 535–568
66 Carpenter, S. and Schumacher, G. A. (1966) *Arch. Neurol.* 14, 82–94
67 Zoghbi, H. Y. *et al.* (1988) *Ann. Neurol.* 23, 580–584
68 Diaz, G. O., Fleites, A. N., Sagaz, C. and Auburger, G. (1990) *Neurology* 40, 1369–1375
69 Landa, R. *et al.* (1992) *Psychol. Med.* 22, 245–254
70 Riggins, G. J. *et al.* (1992) *Nature Genet.* 2, 186–191
71 Wharton, K. A., Yedvobnick, B., Finnerty, V. G. and Artavanis-Tsakonas, S. (1985) *Cell* 40, 55–62
72 Duboule, D., Haenlin, M., Galliot, B. and Mohier, E. (1987) *Mol. Cell. Biol.* 7, 2003–2006
73 Kao, C. C. *et al.* (1990) *Science* 248, 1646–1650
74 Kadonaga, J. T., Courey, A. J., Ladika, J. and Tjian, R. (1988) *Science* 242, 1566–1570
75 Courey, A. J. and Tjian, R. (1988) *Cell* 55, 887–898
76 Haynes, S. R., Rebbert, M. L., Mozer, B. A., Forquignon, F. and Dawid, I. B. (1987) *Proc. Natl Acad. Sci. USA* 84, 1819–1823
77 Mermod, N., O'Neill, E., Kelly, T. J. and Tjian, R. (1989) *Cell* 58, 741–753
78 Ma, J. and Ptashne, M. (1987) *Cell* 48, 847–853
79 He, X. and Rosenfeld, M. G. (1991) *Neuron* 7, 183–196
80 Campeau, S. *et al.* (1991) *Brain Res.* 565, 349–352
81 Cole, A. J., Saffen, D. W., Baraban, J. M. and Worley, P. F. (1989) *Nature* 340, 474–476
82 Guerini, D. and Klee, C. B. (1989) *Proc. Natl Acad. Sci. USA* 86, 9183–9187
83 Edwards, A., Civitello, A., Hammond, H. A. and Caskey, C. T. (1991) *Am. J. Hum. Genet.* 49, 746–756
84 Schalling, M., Hudson, T. J., Buetow, K. H. and Housman, D. E. *Nature Genet.* (in press)
85 Li, S. H., McInnis, M. G., Margolis, R. L., Antonarakis, S. E. and Ross, C. A. (1992) *Am. J. Hum. Gen.* 51, A41
86 Lee, C. C., Panzer, S. W. and Caskey, C. T. (1992) *Am. J. Hum. Genet.* 51, A101
87 Duyao, M. P. *et al. Nature Genet.* (in press)
88 Ross, C. A. *et al. Soc. Neurosci. Abstr.* (in press)
89 Peltomäki, P. *et al.* (1993) *Science* 260, 810–812
90 Aaltonaen, L. E. *et al.* (1993) *Science* 260, 812–816

Erratum

In the article 'The role of positron emission tomography in the assessment of human neurotransplantation' by Guy V. Sawle and Ralph Myers (May 1993, Vol. 16, pp. 172–176), Figures 2 and 3 were transposed relative to their legends. Thus, the coronal section that appears as Figure 3 should appear above the legend for Figure 2 and the parasaggital section that appears as Figure 2 should appear above the legend for Figure 3. We apologise to Drs Sawle and Myers, and to the readers for this error.

Exhibit E

Am. J. Hum. Genet. 54:864-870, 1994

Evidence for Anticipation in Schizophrenia

Anne S. Bassett* and William G. Honer†

*Department of Psychiatry, University of Toronto, and Genetics Section, Schizophrenia Research Program, Queen Street Mental Health Centre, Toronto; and †Department of Psychiatry, University of British Columbia, Vancouver

Summary

Anticipation, or increasing severity of a disorder across successive generations, is a genetic phenomenon with an identified molecular mechanism: expansion of unstable trinucleotide repeat sequences. This study examined anticipation in familial schizophrenia. Three generations of siblines from the affected side of families selected for unilineal, autosomal dominant-like inheritance of schizophrenia were studied (n = 186). Across generations more subjects were hospitalized with psychotic illness ($P<.0001$), at progressively earlier ages ($P<.0001$), and with increasing severity of illness ($P<.0003$). The results indicate that anticipation is present in familial schizophrenia. These findings support both an active search for unstable trinucleotide repeat sequences in schizophrenia and reconsideration of the genetic model used for linkage studies in this disorder.

Introduction

The clinical observation of anticipation—i.e., inherited illness that becomes more severe across successive generations—has recently been found to have a molecular basis: expanding GC-rich trinucleotide repeat sequence mutations (Harper et al. 1992; Sutherland and Richards 1992). In fragile X syndrome (Verkerk et al. 1991), myotonic dystrophy (Fu et al. 1992), spinobulbar muscular atrophy (Brook et al. 1992), spinocerebellar atrophy type 1 (Orr et al. 1993), and Huntington disease (Huntington's Disease Collaborative Research Group 1993), increasing severity of illness, earlier age at onset, and/or increasing proportion of ill individuals in successive generations are associated with longer trinucleotide expansions. Schizophrenia is another neuropsychiatric disorder that may display this anticipation phenomenon and that therefore may have familial forms caused by an unstable trinucleotide repeat.

Schizophrenia is a severe disorder characterized by social withdrawal and psychotic symptoms, such as delusions and hallucinations. The illness has a variable age at onset, often beginning in early adulthood and resulting in lifelong disabilities in social and occupational functioning. Evidence from family, twin, and adoption studies, including those using reliable diagnostic criteria (Lowing et al. 1983; Kendler et al. 1985), strongly supports a genetic etiology for schizophrenia (Gottesman and Shields 1982). However, the mode of inheritance for schizophrenia is not readily identifiable and is proposed to involve interacting genes (Risch 1990). In families with the illness, reduced penetrance and variable expression are commonly found. Other psychotic disorders, of lesser severity, and schizotypal personality traits such as social isolation, odd communication, and extreme suspiciousness are conditions likely reflecting variable expression of genetic susceptibility to schizophrenia (Gottesman and Shields 1982; Lowing et al. 1983; Kendler et al. 1985). These factors, along with the possibility of genetic heterogeneity and the practical difficulties of studying an illness with a significant suicide rate and suspicious, socially isolated individuals, combine to make schizophrenia a challenging disorder for linkage studies (Bassett 1991). Strategies to overcome these difficulties include focusing on familial schizophrenia where inheritance is consistent with Mendelian patterns, using reliable diagnostic methods, highly polymorphic DNA markers, and lod-score methods that model the complexities of the inheritance. Linkage studies to date, using informative fami-

Received September 13, 1993; accepted for publication December 22, 1993.

Address for correspondence and reprints: Dr. Anne S. Bassett, Schizophrenia Research Program, Queen Street Mental Health Centre, 1001 Queen Street West, Toronto, Ontario M6J 1H4, Canada.

0002-9297/94/5405-0016$02.00

lies and models based on Mendelian inheritance, have yielded significant lod scores but no replicated positive results (Bassett 1991).

Dynamic modifications of classical patterns of genetic transmission, such as anticipation (Mott 1911) and genomic imprinting (differential expression of genetic material depending on parental origin of the gene and the underlying molecular mechanisms), may explain the complex genetics of schizophrenia and other major mental illnesses (McInnis et al. 1991; Flint 1992). For example, trinucleotide repeats can cause reduced penetrance and variable expression (Caskey et al. 1992; Sutherland and Richards 1992), by existing in a premutation form, by reductions in repeat size, or by somatic mutation in early embryogenesis (Lavedan et al. 1993). The unstable nature of trinucleotide repeats also provides a possible mechanism for the high mutation rates proposed for schizophrenia (Slater and Cowie 1971). If there were evidence of anticipation in schizophrenia, screening for triplet repeat mutations would become a rational option for gene localization studies, and modification of the genetic model used in linkage studies would need to be considered. The current study investigated whether anticipation was present in a familial schizophrenia sample participating in a linkage study.

Subjects and Methods

Subjects were members of eight extended nonconsanguineous families participating in a genetic linkage study of familial schizophrenia. Local psychiatrists identified prospective pedigrees segregating schizophrenia. Families were selected for large size, availability of two or more generations of adults, and apparent unilineal, autosomal dominant–like inheritance of schizophrenia and genetically related disorders. Bilineal families with evidence, from family or collateral history, of schizophrenia or other nonaffective psychotic disorders on both sides were excluded. Further details of the original ascertainment and assessment for the linkage study are described elsewhere (Bassett et al. 1993). Since families were ascertained in their entirety, proband status could be assigned to all affected subjects. Therefore no subjects were excluded from analyses.

Subjects only from the affected side of each family were taken into account, to determine sibling sets (siblines) in the index generation (IG) (n = 13 siblines), parental generation (PG) (n = 10 siblines), and grandparental generation (GG) (n = 8 siblines). The affected side was defined by (1) a parent hospitalized with psychosis (four cases), (2) an aunt/uncle hospitalized with psychosis (four cases), (3) a parent/aunt/uncle with schizotypal traits (seven cases), or (4) a sibship linking two affected nuclear families (six cases in two extended families) (see fig. 1). In two cases in the GG, the affected side could not be determined, and the smaller of the unaffected maternal or paternal siblines was arbitrarily selected. There were three instances of unknown paternity. In two of these cases, the maternal line was affected. In the third case, neither the mother nor her five siblings were affected, and this GG sibline was not included in the analysis.

Family-history information was obtained for each subject from three or more family members by using the Family History–Research Diagnostic Criteria (FH-RDC) method (Andreasen et al. 1977). Genealogical records were used to confirm dates of birth and death. Medical records were searched back to 1866 for evidence of psychiatric hospitalization. Because the subject families originated and seldom moved from a circumscribed region of Canada, and because the one psychiatric hospital available until the 1980s consistently maintained a comprehensive file-card system of recording admissions, virtually complete ascertainment of psychiatric hospitalization was assured. Records were collected for all subjects with a history of psychiatric hospitalization. Living subjects participating in the linkage study were directly interviewed by a psychiatrist (Bassett et al. 1993). Diagnostic folders containing the family history and, if present, medical records and interview data were reviewed independently by two psychiatrists (A.S.B. and W.G.H.), one of whom (W.G.H.) was blind to the pedigree structures. A consensus lifetime Research Diagnostic Criteria (RDC) diagnosis for psychotic disorders, age at first hospitalization for a psychotic illness, and presence of two or more RDC schizotypal traits were recorded. Psychotic disorders included schizophrenia (n = 25), schizoaffective disorder (n = 13; 12 mainly schizophrenic type and 1 other), unspecified functional psychosis (n = 4), mania with psychosis (n = 1), and depression with psychosis (n = 1). Schizophrenia and schizoaffective disorder were approximately equivalent in severity in the current sample (Bassett et al. 1993) and were considered together in the current study. Individuals with psychotic disorders not severe enough to require hospitalization, as well as subjects with schizotypal traits, were combined in a single schizotypal group, because of small numbers in each category.

Of the 209 subjects in the affected siblines, 23 were excluded from the analyses. One IG subject had not attained the age of 15 years, considered a minimum age

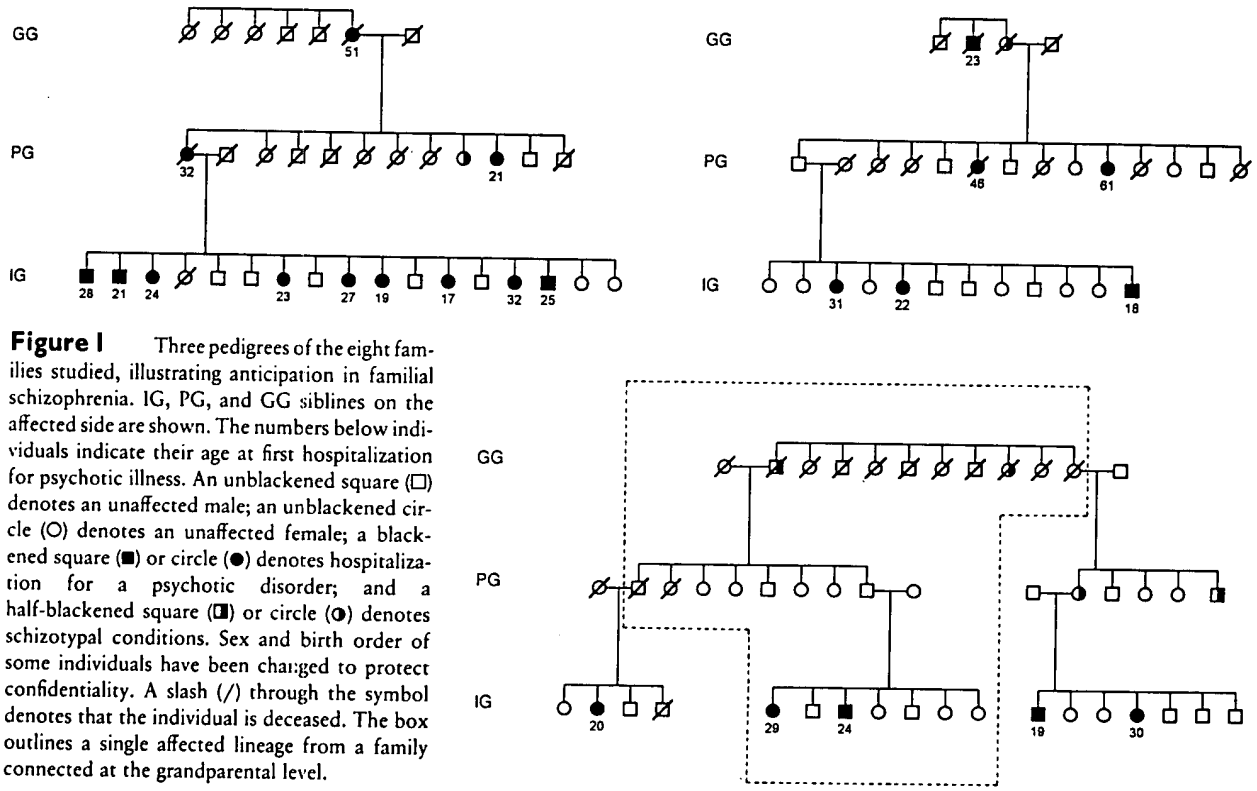


Figure 1 Three pedigrees of the eight families studied, illustrating anticipation in familial schizophrenia. IG, PG, and GG siblines on the affected side are shown. The numbers below individuals indicate their age at first hospitalization for psychotic illness. An unblackened square (□) denotes an unaffected male; an unblackened circle (○) denotes an unaffected female; a blackened square (■) or circle (●) denotes hospitalization for a psychotic disorder; and a half-blackened square (◨) or circle (◑) denotes schizotypal conditions. Sex and birth order of some individuals have been changed to protect confidentiality. A slash (/) through the symbol denotes that the individual is deceased. The box outlines a single affected lineage from a family connected at the grandparental level.

of risk for psychotic illness (Gottesman and Shields 1982). Five subjects (2 PG and 3 GG) had moved, and collateral information was insufficient to determine hospitalization status; and 17 subjects (3 IG, 5 PG, and 9 GG) died before the age of 40 years. Thirteen died in infancy or childhood, two in war, one in an accident at work, and one of unknown cause; none were suicides. Most new cases of schizophrenia may be expected before age 40 years (Gottesman and Shields 1982). Data on the remaining 186 subjects were examined for anticipation, in three ways. First, the rates of hospitalization for psychotic disorders and the rates of schizotypal conditions were compared across generations by using χ^2 analyses. Second, to assess severity of illness, subjects were assigned the following ratings: hospitalized with schizophrenia or schizoaffective disorders—3; hospitalized with other psychotic disorders—2; schizotypal—1; and unaffected—0. Means for each generation were compared using the one-way analysis of variance (ANOVA), including correction for multiple tests of significance with the Student-Newman-Keuls procedure. Third, age at first hospitalization for psychosis was assessed using the life-table method of survival analysis for 1-year intervals. Homogeneity of survival curves over the generations was examined using the Wilcoxon test. The analysis was performed assuming (1) no differential mortality between affected and unaffected and (2) hospitalization rates independent of chronological time (e.g., 1920 vs. 1970). Observations ended at the subject's current age or age at death. Covariates tested were sex and transmission patterns (maternal/paternal).

Results

Demographic characteristics of the sample and results indicating anticipation are presented in table 1. As for other illnesses demonstrating anticipation, expression of illness varied between members of a sibship (fig. 1).

Rates and Distribution of Illness

There were significantly more subjects hospitalized for psychosis across successive generations ($\chi^2 = 16.84$, $P<.0001$, 2 df). Most had schizophrenia or schizoaffective disorders (IG, $n = 30$; PG, $n = 8$; and GG, $n = 0$). Of the six subjects with less severe disorders—unspecified functional psychosis ($n = 4$), psychotic mania ($n = 1$), or psychotic depression ($n = 1$)—four were in PG or GG. Subjects with the least severe illnesses (schizotypal con-

Table 1

Characteristics of the Sample, by Generation (n = 186)

	IG	PG	GG
Total no. of subjects (females)	86 (38)	62 (24)	38 (21)
Mean size of sibline (SD)	6.61 (4.27)	6.20 (3.61)	5.43 (4.20)
Mean age of living[a] (SD)	40.61 (8.74)	65.55 (12.96)	71.80 (8.26)
Mean age at death[b] (SD)	44.25 (3.20)	57.15 (14.58)	77.35 (9.34)
Mean age at first hospitalization (SD)	26.16 (8.28)	34.00 (17.28)	41.33 (15.89)
Subjects hospitalized for psychotic illness	32 (37.21%)	9 (14.52%)	3 (7.89%)
Subjects with schizotypal conditions	8 (9.30%)	8 (12.90%)	8 (21.05%)

[a] No. of living subjects: IG, 82; PG, 42; and GG, 10.

[b] No. of dead subjects with known age at death: IG, 4; PG, 20; and GG, 20. Eight other GG subjects had ages at death that were less precisely known, e.g., "in their 70s." The mean shown did not change significantly when these subjects were included using estimated ages at death.

ditions) were twice as common in the GG as in the IG (21.05% vs. 9.30%); however, across the three generations, the result was not significant ($\chi^2 = 3.24$, $P = .20$, 2 df). When these subjects with schizotypal conditions were included with hospitalized subjects, there were still significantly more affected subjects in the youngest generation ($\chi^2 = 6.86$, $P = .032$, 2 df).

Because of the possibility of compounded error with families being connected at the parental or grandparental level, analyses were rerun with a single affected lineage from each of the seven kindreds that had data from all three generations, with the largest siblines being selected (see fig. 1). The same results were found using this subsample of 151 subjects, for both hospitalization rates ($\chi^2 = 12.30$, $P<.002$, 2 df) and hospitalization plus schizotypal rates ($\chi^2 = 9.46$, $P<.009$, 2 df). Excluding subjects who died before age 40 or moved away (n = 22) could have influenced the study's findings, since most were in the two senior generations. When results with these subjects and with penetrance estimated maximally at 100% were considered, so that one-half ($n = 11$) would have been hospitalized with psychotic disorders, the results for hospitalization rates would still have remained significant ($\chi^2 = 10.02$, $P<.01$, 2 df). Data were also reanalyzed examining the effect that increasing hospitalization rates over time would have on the results. Secular trends of hospitalization rates for specific illnesses were not available. Therefore, arbitrary increases were tested, by adding 100% more subjects ($n = 3$) to the hospitalized GG group and 50% more ($n = 4.5$) to the PG group. Under these conditions, a significant increase in the rate of psychosis requiring hospitalization would continue to be present over the generations ($\chi^2 = 7.75$, $P<.05$, 2 df).

Severity of Illness

Means for four-point severity-of-illness ratings for the three generations were as follows: IG ($n = 86$), 1.19; PG ($n = 62$), 0.58; and GG ($n = 38$), 0.37. Severity of illness significantly increased over the generations ($F = 8.39$, 2 df, $P = .0003$). Pairwise comparisons using the Student-Newman-Keuls test revealed that the significant difference was between the IG and the PG and GG. PG and GG severity means were not significantly different from each other.

Age at First Hospitalization

The survival curve (fig. 2) shows that subjects were first hospitalized for psychosis at progressively younger ages, across generations ($\chi^2 = 26.76$, $P = .0001$, 2 df). Sex of subject was not a significant covariate to generation, for age at first hospitalization (increment $\chi^2 = 0.40$, $P = .53$, 1 df). All of the GG and most (68%) of the PG subjects had achieved an age of 55 years or more, by which time they would have been virtually through the age at risk for schizophrenia (Gottesman and Shields 1982).

Imprinting

There were equal rates of maternal and paternal transmission (six and seven cases, respectively) from the PG to the IG. The mean age at first hospitalization for maternal transmission (PG to IG) was 25.00 years (SD 5.94 years), and that for paternal transmission was 27.64 years (SD 10.63 years), a nonsignificant difference ($t = .83$, $P = .41$). Grandparental to parental transmission was predominantly maternal (seven of nine cases), with one unknown and only one example of paternal transmission. Sex of transmitting parent was not a sig-

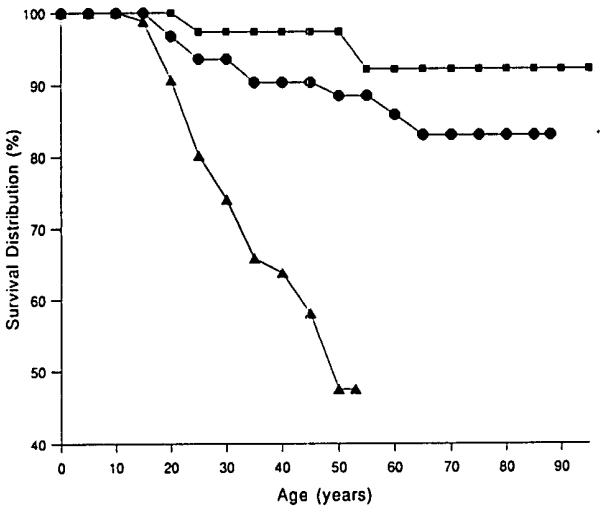


Figure 2 Survival curves for age at first hospitalization for psychotic illness, comparing IG (▲), PG (●), and GG (■). Annual percent surviving without hospitalization is plotted for every 5th year. Observation of individuals ended at their age of death or, if they were living, at their current age.

nificant covariate in the survival analysis examining age at first hospitalization (increment $\chi^2 = 0.32$, $P = .57$, 2 df).

Discussion

The results suggest that familial schizophrenia exhibits anticipation. All of the families studied showed this phenomenon, manifest as increasing rates of hospitalized psychotic illness, worsening severity of illness, and/or earlier age at onset, across successive generations (fig. 1). These findings are consistent with differences in rates of hospitalization and age-at-onset data for parent-child pairs in studies of schizophrenia over the century (Mott 1910; Kay 1963; Penrose 1971; Decina et al. 1991). In each of these studies, rates of hospitalization for psychosis were lower for antecedent generations, and age at onset for parents was significantly later than that for offspring. Investigations of ancestors and extended families also support these findings (Karlsson 1966; Odegaard 1972; Wetterberg and Farmer 1991). As well, less severe psychotic illnesses (e.g., affective disorders) are consistently more common in the generation antecedent to schizophrenic probands (Slater and Cowie 1971; Bleuler 1978). These results complement reported morbid risk of schizophrenia for parents, which is almost half that for siblings (Gottesman and Shields 1982). While alternative reasons, including selection biases such as reduced fertility in earlier-onset schizophrenia, have been proposed to explain these clinical observations, they are all consistent with the phenomenon of anticipation in schizophrenia.

In contrast, findings from the current study are only suggestive of sex-specific differences in the transmission of schizophrenia. However, the possibility of an excess maternal over paternal transmission in schizophrenia is consistent with trends found both recently by others (Sharma et al. 1993) and in studies of large data sets in the older literature (Penrose 1971; Slater and Cowie 1971). Fertility may be especially low in male patients with schizophrenia (Gottesman and Shields 1982), and in the current study this could be the reason for the high rate of maternal transmission from the GG lines to the PG lines. If this were the case, however, one would have expected predominantly maternal inheritance from the PG to the IG lines; but maternal and paternal rates were equal. Comparable transmission patterns have been found to be due to greater variation of trinucleotide repeat length after female meiosis in myotonic dystrophy (Lavedan et al. 1993). The effect may be subtle, requiring larger samples to demonstrate imprinting in schizophrenia.

There are several possible biases that can explain results that indicate anticipation (Penrose 1948). First, subjects could have died before expressing the mutation. However, all of the GG and most of the PG lived beyond age 55 years. Also, significantly different hospitalization rates across generations remained in the current study, even when half of those who had moved or died before age 40 years were assigned affected status. Second, reduced fertility of individuals with earlier onset of schizophrenia could cause preferential ascertainment of parents with later onset. This bias should be minimized in the current study, because (1) few parents were affected with psychosis and (2) siblines were large in all three generations, providing multiple opportunities for detection of affecteds, regardless of their fertility. Third, subjects in the IG could have been too young to yet express a late-onset form of psychosis, which would attenuate the age-at-onset findings. Since one-half of the subjects in the IG were over age 40 years, most were beyond the period of highest risk. Even if more new cases of psychosis subsequently arose, this would most likely occur in the IG and would only serve to strengthen the present study's findings with respect to differential rates of illness.

In contrast to major depression (Gershon et al. 1987), there is no evidence for a cohort effect in schizophrenia. However, secular trends, such as improved detec-

tion, that could, over time, lead to higher hospitalization rates and/or younger age at first hospitalization for psychosis are important to consider, since these could have influenced the principal findings of the current study. In the literature (Mott 1910; Kay 1963; Penrose 1971; Decinà et al. 1991), examination of ages at first hospitalization did not reveal secular trends to younger age over the century. In the current study, arbitrary increases in hospitalization incidence assigned to the PG or GG did not change the observation of anticipation. Specific factors that may influence secular trends in hospitalization, including drug abuse and psychosocial stressors, do not appear to have played a role in the sample studied. Only three hospitalized subjects in the IG had a history of stimulant or hallucinogen use, drugs that may in some cases precipitate a psychotic illness. On the basis of direct interviews, it appeared that psychosocial stressors endured by PG and GG, such as the World Wars and Great Depression, were more severe than those faced by the IG.

Another possibility is that the family-history method tends to underestimate rates of psychiatric disturbance, particularly in relatives who are dead or less known (Andreasen et al. 1977). The consequence could be that actual rates for schizotypal traits could be higher than those found, particularly for PG and GG. However, results for severe illness requiring hospitalization would remain unchanged. Another factor that could have compounded errors in the PG and GG was the use of IG siblines connected at the parental or grandparental level. However, both the fact that results remained the same when only one ascending line from each kindred was examined and the fact that at least 13 affected siblines had resulted from eight originating GG lines support the finding of anticipation.

A limitation that is important to consider in studies of common illnesses is the potential for assortative mating to cause an apparently increased prevalence of illness in offspring. This possibility was minimized by selecting unilineal pedigrees with no evidence of schizophrenia or related disorders in the married-in person, their siblings, or parents. Although individuals marrying in could have been nonexpressing carriers of the disorder with nonexpressing close relatives, the likelihood appears small that assortative mating could account for the results in the current study. Because the families studied were selected because of their autosomal dominant–like inheritance and large sibships, the results may not be generalizable to schizophrenia in the general population, although they are consistent with observations from large population-based samples (Penrose 1971; Bleuler 1978). Also, the schizophrenia in the subject families may be a particularly severe form, which could exaggerate the findings. However, the mean age at onset for the IG is similar to others' results (Mott 1911; Gottesman and Shields 1982; Decina et al. 1991; Sharma et al. 1993). Data on specific symptom patterns suggest that the familial schizophrenia in the present families is comparable in nature and in severity to samples drawn from the general population (Bassett et al. 1993).

Despite the possible biases and limitations, the weight of evidence from both the current investigation and the literature is consistent with the finding of anticipation in familial schizophrenia. Other families with schizophrenia should be examined for anticipation and possible accompanying maternal imprinting phenomena, to confirm the current study's findings. Although published pedigrees consistently show evidence of anticipation (Karlsson 1966; Wetterberg and Farmer 1991), complete ascertainment of other large kindreds with contemporary reliable diagnostic assessments would be useful. However, the most exciting possibilities—and the confirmation of the clinical observations of the current study—lie in the search for expanding trinucleotide repeats and other DNA sequence mutations in schizophrenia. New methods becoming available to detect these mutations (Orr et al. 1993; Schalling et al. 1993) will complement and may accelerate the search for pathological genes in linkage studies. In addition, the current study has implications for the genetic model used in linkage studies. Parameters, particularly penetrance, which would vary according to generation, may need to be modified to reflect the effects of anticipation and possibly imprinting. In the light of clinical evidence for anticipation, these strategies represent real promise for deciphering the genetics of schizophrenia.

Acknowledgments

The authors are grateful to the families for their cooperation; to J. McAlduff, R.N., and A. Bury, B.A., for invaluable assistance; and to N. Risch, Ph.D., for comments on the manuscript. This work was supported in part by funding from the Scottish Rite Schizophrenia Research Program, Ian Douglas Bebensee Foundation, Ontario Mental Health Foundation, and Medical Research Council of Canada.

References

Andreasen N, Endicott J, Spitzer R, Winokur G (1977) The family history method using diagnostic criteria: reliability and validity. Arch Gen Psychiatry 34:1228–1235

Bassett AS (1991) Linkage analysis of schizophrenia: challenges and promise. Soc Biol 38:189–196

Bassett AS, Collins EJ, Nuttall S, Honer WG (1993) Positive and negative symptoms in families with schizophrenia. Schizophr Res 11:9–19

Bleuler M (1978) The schizophrenic disorders: long-term patient and family studies. Yale University Press, New Haven

Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, et al (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. Cell 68:799–808

Caskey CT, Pizzuti A, Fu Y, Fenwick RG Jr, Nelson DL (1992) Triplet repeat mutations in human disease. Science 256:784–788

Decina P, Mukherjee S, Lucas L, Linder J, Horwath E (1991) Patterns of illness in parent-child pairs both hospitalized for either schizophrenia or a major mood disorder. Psychiatry Res 39:81–87

Flint J (1992) Implications of genomic imprinting for psychiatric genetics. Psychol Med 22:5–10

Fu Y-H, Pizzuti A, Fenwick RG Jr, King J, Rajnarayan S, Dunne PW, Dubel J, et al (1992) An unstable triplet repeat in a gene related to myotonic muscular dystrophy. Science 255:1256–1258

Gershon ES, Hamovit JH, Guroff JJ, Nurnberger JI (1987) Birth-cohort changes in manic and depressive disorders in relatives of bipolar and schizoaffective patients. Arch Gen Psychiatry 44:314–319

Gottesman II, Shields J (1982) Schizophrenia: the epigenetic puzzle. Cambridge University Press, Cambridge

Harper PS, Harley HG, Reardon W, Shaw DJ (1992) Anticipation in myotonic dystrophy: new light on an old problem. Am J Hum Genet 51:10–16

Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72:971–983

Karlsson JL (1966) The biologic basis of schizophrenia. Charles C Thomas, Springfield, IL

Kay DWK (1963) Late paraphrenia and its bearing on the aetiology of schizophrenia. Acta Psychiatr Scand 39:159–169

Kendler KS, Gruenberg AM, Tsuang MT (1985) Psychiatric illness in first-degree relatives of schizophrenic and surgical control patients. Arch Gen Psychiatry 42:770–779

Lavedan C, Hofmann-Radvanyi H, Shelbourne P, Rabes J-P, Duros C, Savoy D, Dehaupas I, et al (1993) Myotonic dystrophy: size- and sex-dependent dynamics of CTG meiotic instability, and somatic mosaicism. Am J Hum Genet 52:875–883

Lowing PA, Mirsky AF, Pereira R (1983) The inheritance of schizophrenia spectrum disorders: a reanalysis of the Danish adoption study data. Am J Psychiatry 140:1167–1171

McInnis MG, McMahon FJ, Chase GA, Simpson SG, Ross CA, DePaulo JR Jr (1991) Anticipation in bipolar affective disorder. Am J Hum Genet 53:385–390

Mott FW (1910) Hereditary aspects of nervous and mental diseases. Br Med J 2:1013–1020

——— (1911) Heredity and insanity. Lancet 3:1251–1259

Odegaard O (1972) The multifactorial theory of inheritance in predisposition to schizophrenia. In: Kaplan AR (ed) Genetic factors in schizophrenia. Charles C Thomas, Springfield, IL, pp 256–275

Orr HT, Chung M, Banfi S, Kwiatkowski TJ Jr, Servadio A, Beaudet AL, McCall AE, et al (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nature Genet 4:221–226

Penrose LS (1948) The problem of anticipation in pedigrees of dystrophia myotonica. Ann Eugenics 14:125–132

——— (1971) Critical survey of schizophrenia genetics. In: Howells JG (ed) Modern perspectives in world psychiatry. Oliver & Boyd, Edinburgh, pp 3–19

Risch N (1990) Linkage strategies for genetically complex traits. I. Multilocus traits. Am J Hum Genet 46:222–228

Schalling M, Hudson TJ, Buetow KH, Housman DE (1993) Direct detection of novel expanded trinucleotide repeats in the human genome. Nature Genet 4:135–139

Sharma T, Brynjolfsson J, Butler R, Read T, Murphy P, Sigmundsson T, Petursson H, et al (1993) Genomic imprinting in the Icelandic/British schizophrenia study: preliminary results. Schizophr Res 9:124–125

Slater E, Cowie V (1971) The genetics of mental disorders. Oxford University Press, London

Sutherland GR, Richards RI (1992) Anticipation legitimized: unstable DNA to the rescue. Am J Hum Genet 51:7–9

Verkerk AJMH, Pieretti M, Sutcliffe JS, Fu Y-H, Kuhl DPA, Pizzuti A, Reiner O, et al (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 65:905–914

Wetterberg L, Farmer AE (1991) Clinical polydiagnostic studies in a large Swedish pedigree with schizophrenia. Eur Arch Psychiatry Clin Neurosci 240:188–190

Exhibit F

Am. J. Hum. Genet. 53:385–390, 1993

Anticipation in Bipolar Affective Disorder

Melvin G. McInnis, Francis J. McMahon, Gary A. Chase, Sylvia G. Simpson, Christopher A. Ross, and J. Raymond DePaulo, Jr.

Department of Psychiatry and Behavioral Sciences, The Johns Hopkins University School of Medicine, Baltimore

Summary

Anticipation refers to the increase in disease severity or decrease in age at onset in succeeding generations. This phenomenon, formerly ascribed to observation biases, correlates with the expansion of trinucleotide repeat sequences (TNRs) in some disorders. If present in bipolar affective disorder (BPAD), anticipation could provide clues to its genetic etiology. We compared age at onset and disease severity between two generations of 34 unilineal families ascertained for a genetic linkage study of BPAD. Life-table analyses showed a significant decrease in survival to first mania or depression from the first to the second generation ($P < .001$). Intergenerational pairwise comparisons showed both a significantly earlier age at onset ($P < .001$) and a significantly increased disease severity ($P < .001$) in the second generation. This difference was significant under each of four data-sampling schemes which excluded probands in the second generation. The second generation experienced onset 8.9–13.5 years earlier and illness 1.8–3.4 times more severe than did the first generation. In additional analyses, drug abuse, deaths of affected individuals prior to interview, decreased fertility, censoring of age at onset, and the cohort effect did not affect our results. We conclude that genetic anticipation occurs in this sample of unilineal BPAD families. These findings may implicate genes with expanding TNRs in the genetic etiology of BPAD.

Introduction

The term "anticipation" describes an inheritance pattern within a pedigree wherein disease severity increases or age at onset decreases in successive generations (Howeler et al. 1989; Harper et al. 1992). Early in this century, anticipation was reported in myotonic dystrophy (DM) (Fleischer 1918) but was later discounted as a bias of ascertainment (Penrose 1948). Recent clinical investigations have supported the presence of anticipation in DM (Howeler et al. 1989), noting that it was likely to be inherent in the mode of transmission. Anticipation has also been shown in fragile X syndrome and helps explain the apparent non-Mendelian features of its inheritance (the Sherman paradox; Sherman et al. [1985]). Recently, anticipation in DM (Mahadevan et al. 1992) and fragile X (Fu et al. 1991) was directly associated with expansion of trinucleotide repeats (TNRs) at the disease locus (Caskey et al. 1992).

Some early observers claimed that the clinical phenomenon of anticipation was present in the inheritance of psychiatric disorders (Morel 1857; Mott 1910, 1911). However, no studies have addressed anticipation in psychiatric disorders since Kraepelin (1921) separated schizophrenia from bipolar affective disorder (BPAD) early in this century. BPAD family (Winokur et al. 1982; Rice et al. 1987), twin (Bertelsen et al. 1977), and adoption studies (Mendlewicz and Rainer 1977) demonstrate genetic transmission, but the mode of inheritance remains uncertain (Rice et al. 1987). Anticipation could be one of the factors influencing the mode of transmission.

In view of (1) the new data describing the molecular basis of anticipation and (2) the persistent uncertainty about the mode of inheritance of BPAD, we examined 34 families ascertained for a genetic linkage study of BPAD for evidence of anticipation. Our findings sug-

Received January 6, 1993; revision received March 18, 1993.

Address for correspondence and reprints: Melvin G. McInnis, Department of Psychiatry and Behavioral Sciences, The Johns Hopkins University School of Medicine, 600 North Wolfe Street, Baltimore, MD 21287-7463.

0002-9297/93/5302-0010$02.00

gest that anticipation occurs in this sample of families, as demonstrated by an earlier age at onset and greater severity of illness in successive generations.

Subjects and Methods

Ascertainment

Subject selection and evaluation for the genetic linkage study of BPAD are described in detail elsewhere (Simpson et al. 1992) and will be summarized here. Clinics and inpatient units in Baltimore and Iowa City have been screened for treated bipolar I (BPI) or bipolar II (BPII) probands with two or more sibs or with one sib and one parent with an affected phenotype by family history. BPI disorder is defined by the presence of "mania," a period of severely elevated mood and activity. In BPII disorder the periods of elevated mood or activity are less severe and, while noticeable to the patient or family, are not incapacitating (Coryell et al. 1984). For the purposes of ascertainment, the affected phenotype included any bipolar, unipolar depressive, or schizoaffective disorder. Eighteen hundred probands have been screened. After a multi-informant family psychiatric history was obtained, probands with an affected phenotype in both parental lines and probands whose key relatives were unavailable or uncooperative were excluded, leaving 125 families meeting the ascertainment criteria.

Evaluation

In these 125 families, probands and all available first- and second-degree relatives were interviewed by a psychiatrist using the Schedule for Affective Disorders and Schizophrenia—Lifetime version (SADS-L) (Endicott and Spitzer 1978). All diagnoses conformed to the Research Diagnostic Criteria (RDC) (Spitzer et al. 1975). Interviews have been completed on 599 subjects. For the anticipation analysis, the affected phenotype included (1) major depression, single episode (MDS), (2) major depression, recurrent (MDR), (3) BPII, with recurrent major depression, (4) BPI, and (5) schizoaffective disorder, manic type (SA).

Study Sample

For the present study we excluded from the 125 ascertained families (1) families with clinical evidence of bilineality, on the basis of either the presence of an affected phenotype or even a single episode of major depression in both parental lines (the additive effect of disease genes from both parental lines could create the appearance of anticipation); and (2) families without at least one interviewed, affected individual in each of two successive generations (necessary for the comparison of affected individuals in successive generations). This left 34 families in the present study.

The older generation (G1)—which usually consisted of parents, aunts, and uncles of the proband—included 63 affected and 78 unaffected individuals. The younger generation (G2), which usually consisted of sibs and first cousins of the proband, included 107 affected and 20 unaffected individuals. We did our analysis under the assumption of autosomal dominant transmission of the affected phenotype. The ratio of affected to unaffected G2 individuals (107:20) represents only those family members directly interviewed by a psychiatrist and includes the 30 G2 probands. In four families the proband and his or her sibs constituted G1, while the proband's children, nieces, and nephews constituted G2. We obtained extensive family history information on all members of the G2 generation, regardless of whether they were interviewed, but we did not include the uninterviewed relatives in the anticipation analysis. When we use family history and direct interview data to diagnose the relatives in the G2 generation (excluding probands), the ratio of affected to unaffected individuals (74:57) does not differ significantly from the expected autosomal dominant ratio ($\chi^2 = 2.18$, $P > .10$).

Anticipation Measures

The measures of anticipation were (1) age at onset and (2) episode frequency. The age at onset was defined as the first episode of RDC major depression or mania (whichever was earlier). These events are assessed, with high reliability, by the SADS-L (Egeland et al. 1987). Episode frequency, considered a measure of disease severity, was calculated on the basis of the total number of episodes of mania and depression divided by the total number of years from onset to interview (age at interview minus age at first mania or depression).

Sampling Schemes

A difficult statistical problem in family studies is that relatives do not constitute strictly independent data points. For example, their onset ages for a given genetic disorder are often more closely correlated than onset ages of nonrelatives. Since there is no simple solution to this problem, we performed analyses on four different samples of pairs selected from our families. Each sampling scheme is subject to different criticisms, but, if all schemes lead to the same conclusions, the results are unlikely to be the consequence of weaknesses of any one sampling scheme.

In all sampling schemes, probands were excluded when part of G2. This avoids any bias arising from the ascertainment of probands when their illness became severe enough to require treatment.

The four sampling schemes were as follows:

I. Random pairs.—One randomly selected affected relative from G1 was paired with one randomly selected affected relative from G2 in each family, to form a random pair (n = 34 pairs). This is a statistically conservative scheme, but it fails to use all affected individuals. By including childless members of G1, this scheme, as well as scheme III below, addresses the bias arising from the possibility that severely ill members of G1 will not produce children.

II. Random transmitting pairs.—One randomly selected affected parent was paired with one randomly selected affected child. In four families there was no affected parent. For these families only, the unaffected parent was arbitrarily assigned an onset age equal to the age at interview and an episode frequency of zero. Thirty-four randomly selected transmitting pairs were generated. Although this statistically conservative scheme fails to include childless members of G1, it is the most direct test of the hypothesis that the illness is increasing in severity when passed from parents to their children.

III. All possible pairs.—Each affected member of G1 was matched with each affected member of G2, to generate several pairs per family, with a total of 196 possible intergenerational pairs. Although this scheme uses some data points more than once, it tests the robustness of the findings based on the random pairs. It also includes all affected members of G1, even if they produced no children.

IV. All possible transmitting pairs.—Each affected parent in G1 was matched with each of his or her affected children, to generate 97 pairs. Similarly to scheme III, this scheme uses some data points more than once but tests the robustness of the findings based on the random transmitting pairs.

Statistics

Onset age was compared between all members of G1 and all members of G2 by using life-table analysis and the Gehan's generalized Wilcoxon test. Pairwise intergenerational differences were compared by using the nonparametric Wilcoxon matched-pairs statistic, since onset age and illness frequency all showed highly non-Gaussian distributions. The Cox proportional hazards model was used to analyze the relationship between generation and onset age while controlling for the effect of birth cohort.

All analyses were performed by using CSS-Statistica (Statsoft). The Cox model was analyzed using PECAN, version 2.4 (Wisconsin Clinical Cancer Center).

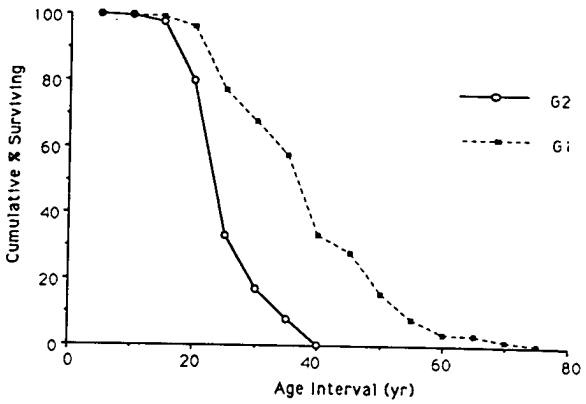


Figure 1 Cumulative survival to first mania or major depression of affected individuals in the older generation (G1) compared with the younger generation (G2).

Results

Life-Table Analysis of Onset Age

Figure 1 shows survival to first mania or major depression for members of G1 versus members of G2. As a group, G2 experienced a significantly earlier onset of illness compared with G1 (Gehan Wilcoxon test statistic = −5.75388, $P < .001$). Fifty percent of affected subjects in G2 became ill by age 23 years, as compared with age 37 years for members of G1, a difference of 14 years.

Pairwise Comparisons

In pairwise comparisons, members of G2 experienced a significantly earlier illness onset than did members of G1 (table 1). This significantly earlier onset for G2 was observed under all four sampling schemes. Members of G2 fell ill 8.9–13.5 years earlier than did members of G1. Overall, 79% of random pairs, 79% of random transmitting pairs, 83% of all possible pairs, and 80% of all possible transmitting pairs showed anticipation based on age at onset.

Table 1 also shows that members of G2 had a significantly higher frequency of episodes than did members of G1. Again, a significant difference was seen regardless of sampling scheme. Members of G2 experienced an episode of illness 1.8–3.4 times as often as did

Table 1

Pairwise Comparisons

	Sampling Scheme			
	I	II	III	IV
Median age at onset:				
G1	30	35	30	30
G2	18	18	17	19
Median difference	12	12.5	13.5	8.9
Z^a	4.4	4.2	10.6	6.9
P	<.001	<.001	<.001	<.001
Median episode frequency:				
G1	.17	.16	.20	.16
G2	.31	.31	.62	.55
Median difference:	.15	.17	.32	.3
Z^a	3.3	2.84	6.76	5.88
P	<.001	<.001	<.001	<.001

[a] Wilcoxon matched-pairs test.

members of G1. Anticipation based on episode frequency was seen in 76% of random pairs, 76% of random transmitting pairs, 75% of all possible pairs, and 79% of all possible transmitting pairs.

Changes in their age at onset or episode frequency which were consistent with anticipation were seen in 95.3% of all pairs. There was no significant effect of the sex of the affected parent on the intergenerational differences in either age at onset ($P > .16$) or episode frequency ($P > .27$).

In the course of data analysis it was observed that, in 26 of 34 random transmitting pairs, there was a change in the form of the affected phenotype (e.g., from an MDR parent to a BPI offspring). In 18 of these pairs either the diagnosis of unaffected, MDS, or MDR in the parent became a diagnosis of BPII or BPI in the offspring or there was a shift from BPII in the parent to BPI in the offspring. In the remaining eight pairs the reverse occurred. When subjected to the sign test, this trend toward "phenotypic progression" did not reach significance ($P = .08$).

Correction for Possible Biases

We attempted to correct for several possible biases in post hoc analyses of the random pairs. Since substance abusers may have an earlier onset and more severe affective illness, the analyses were repeated after the five pairs containing a substance abuser in G2 were eliminated. All differences remained significant ($P < .01$).

Another potential bias arises from the possibility that severely affected individuals die young and thus are less likely to be available for evaluation as members of G1. We determined that 5 of 34 deceased members of G1 were likely to have been affected, on the basis of family history from multiple informants. We assigned these individuals an onset age of 19 years and an episode frequency score of .315, the median values for members of G2, and repeated the analyses. All differences remained significant ($P < .001$).

Since subjects cannot be observed to fall ill later than the age at which they were examined, G2 may have a paucity of late-onset cases simply because members of G2 are younger, thus creating a false impression of anticipation. In an attempt to address this bias, we assigned the 20 unaffected members of G2 an onset age of 61.7 years, equal to the median age of contact for G1. Despite this correction, onset age remained significantly lower in G2 versus G1 ($P < .001$).

Finally, we addressed the cohort effect (Gershon et al. 1987), which concerns the reported earlier onset of major affective disorder for members of cohorts born since 1945. This phenomenon could create the appearance of anticipation, since members of G2 tended to belong to later birth cohorts. To address this bias, we performed a bivariate analysis using the Cox proportional hazards model (Cox 1972), with birth cohort and generational membership as covariates and with age at first mania or major depression as the endpoint. Subjects were divided into three birth cohorts: (1) those born before 1945 (58 in G1, 11 in G2); (2) those born between 1945 and 1960 (5 in G1, 72 in G2); and (3) those born after 1960 (0 in G1, 24 in G2). The results indicated a significant correlation of generational membership with age at onset ($\beta = 0.7658$, $Z = 2.212$, $P = .027$) even when a significant effect of birth cohort was controlled for (cohort 2 vs. cohort 1: $\beta = 0.8355$, $Z = 3.493$, $P = .013$; cohort 3 vs. cohort 1: $\beta = 2.02$, $Z = 4.522$, $P < .001$).

Discussion

In this study, we report data supporting the occurrence of anticipation in 34 unilineal bipolar families. Under each of four different data-sampling schemes, members of the younger generation manifested a significantly more severe form of illness, on the basis of age at onset and episode frequency. This phenomenon was observed in the majority of our families, with 95% of families showing anticipation based on either onset age or episode frequency.

The appearance of anticipation can be created by

several different observation biases. We attempted to address these biases in our study design and in post hoc analyses. The study design was based on two measures of anticipation and on four schemes for sampling the data. Thus it is not likely that our findings are due to the particular weaknesses of any one measure or sampling scheme. Further, the study design sought to eliminate bilineal families, thus minimizing the possible additive effect of disease genes from both parental lines converging in the younger generation. Finally, the study design incorporated sampling schemes which included childless members of the older generation, thus minimizing the bias of severely ill individuals having fewer children.

All results remained significant when the possible impact of substance abuse, differential mortality among members of the older generation, censoring of onset age, and the cohort effect were addressed in post hoc analyses. Thus we conclude that our findings are not artifacts arising from these observational biases.

In an attempt to control for the effect of an individual's being raised by an affected parent, we examined our data by using sampling scheme II (random transmitting pairs). We found that, in 14 of the 34 families, the G2 individual had onset of illness prior to that of the G1 parent. In these families there was significant evidence of anticipation, on the basis of age at onset and episode frequency. This suggests that a genetic effect underlying our observation is likely, although a more definitive answer to this issue might be provided by adoption studies.

Two further biases deserve mention. Older subjects may report a later than actual age at onset because they forget earlier episodes (Rice et al. 1987; Warshaw et al. 1991). We addressed but did not wholly overcome this problem by collecting data in a semistructured interview with a psychiatrist, wherein subjects' reports were exposed to medical cross-examination. If our findings were due primarily to forgetting, we would expect a smaller generational difference for events less likely to be forgotten. However, when we analyzed age at first mania, defined by RDC as an "incapacitating event" (and thus less likely to be forgotten), there was an even larger generational difference in onset age (G1 = 40 years of age, G2 = 21 years of age; $P < .001$).

Our ascertainment scheme, which required three affected members (including proband) per family, may have introduced some bias into the analysis. Such families are not likely to be representative of families affected by bipolar disorder in the community. They are likely to be enriched for genetic cases. There may also be an overrepresentation of severely ill individuals in G2, because of a tendency to preferentially recruit families that exceed the ascertainment criteria. Simply removing the proband from analyses would not overcome this bias. However, we did not require that affected sibs of the proband meet any severity criterion; indeed, 80% of the affected sibs were never hospitalized for their affective disorder, and 46% never received treatment.

Finally, there is the problem, first described by Penrose (1948), that pairs consisting of a parent with early-onset illness and a child with late-onset illness are unlikely to be ascertained by any study, since there is such a large span of time separating the two onset events. The effect of this bias is difficult to measure. However, we expect that the effect would be small in our sample. First, we have ascertained some early-onset parents; in 24% of our families, the affected parent had an onset of illness at $\leq$24 years of age. Second, we have already observed most of the illness onsets that are likely to occur in the second generation. In this generation 70% (excluding probands) were affected at interview, including 13 subjects with onset ages $\geq$30 years of age. Thus it is unlikely that the Penrose bias is exerting a large effect on this sample.

We conclude that anticipation appears to occur in our sample of families with BPAD. These preliminary findings are of potential importance, since the presence of anticipation in even a subset of families with BPAD may explain in part the complex pattern of inheritance of the disorder, including its variable penetrance and expressivity. Our observations need to be confirmed in a group of families ascertained without regard to family history. If confirmed, our findings may implicate genes containing expanding TNRs in the etiology of BPAD.

Acknowledgments

This work was supported in part by NIMH grant MH 42243 (to J.R.D.), by NIMH training grant MH 15330-14 (to M.G.M. and F.J.M.), and by NARSAD Young Investigator awards (to M.G.M. and S.G.S.). We gratefully acknowledge the critical reviews by Deborah Meyers, Colin Stine, Susan Folstein, and David Botstein.

References

Bertelsen A, Harvald B, Hauge M (1977) A Danish twin study of manic depressive disorders. Br J Psychiatry 130:330–351

Caskey CT, Pizzuti A, Fu Y-H, Fenwick RG, Nelson DL

(1992) Triplet repeat mutations in human disease. Science 256:784–789

Coryell W, Endicott J, Reich T, Andreasen N, Keller M (1984) A family study of bipolar II disorder. Br J Psychiatry 145:49–54

Cox DR (1972) Regression models and life tables. J R Soc Stat 34:187–220

Egeland JA, Blumenthal RL, Nee J, Sharpe L, Endicott J (1987) Reliability and relationship of various ages of onset criteria for major affective disorder. J Affect Disord 12:159–165

Endicott J, Spitzer RL (1978) A diagnostic interview: the schedule for affective disorders and schizophrenia. Arch Gen Psychiatry 35:837–844

Fleischer B (1918) Über myotonische Dystrophie mit Katarakt: eine hereditäre, familiäre Degeneration. Arch Ophthalmol 96:91–133

Fu YH, Kuhl DP, Pizzuti A, Pieretti M, Sutcliffe JS, Richards S, Verkerk AJ, et al (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. Cell 67:1047–1058

Gershon ES, Hamovit JH, Guroff JJ, Nurnberger JI (1987) Birth-cohort changes in manic and depressive disorders in relatives of bipolar and schizoaffective patients. Arch Gen Psychiatry 44:314–319

Harper PS, Harley HG, Reardon W, Shaw DJ (1992) Anticipation in myotonic dystrophy: new light on an old problem. Am J Hum Genet 51:10–16

Howeler CJ, Busch HFM, Geraedts JPM, Niermeijer MF, Staal A (1989) Anticipation in myotonic dystrophy: fact or fiction? Brain 112:779–797

Kraepelin E (1921) Manic-depressive insanity and paranoia. E & S Livingstone, Edinburgh

Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, Neville C, et al (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3′ untranslated region of the gene. Science 255:1253–1255

Mendlewicz J, Rainer JD (1977) Adoption study supporting genetic transmission in manic-depressive illness. Nature 268:327–329

Morel BA (1857) Traite des degenerescences. JB Bailliere, Paris

Mott FW (1910) Hereditary aspects of nervous and mental diseases. Br Med J 2:1013–1020

——— (1911) A lecture on heredity and insanity. Lancet 1:1251–1259

Penrose LS (1948) The problem of anticipation in pedigrees of dystrophia myotonica. Ann Eugenics 14:125–132

Rice J, Reich T, Andreasen NC, Endicott J, Eerdewegh MV, Fishman R, Hirschfeld RMA, et al (1987) The familial transmission of bipolar illness. Arch Gen Psychiatry 44:441–447

Sherman SL, Jacobs PA, Morton NE, Froster-Iskenius U, Howard-Peebles PN, Nielsen KB, Partington NW, et al (1985) Further segregation analysis of the fragile X syndrome with special reference to transmitting males. Hum Genet 69:3289–3299

Simpson SG, Folstein SE, Meyers DA, DePaulo JR (1992) Assessment of lineality in bipolar I linkage studies. Am J Psychiatry 149:1660–1665

Spitzer RL, Endicott J, Robins E (1975) Research diagnostic criteria for a selected group of functional disorders. Biometrics Research, New York

Warshaw MG, Klerman GJ, Lavori PW (1991) Are secular trends in major depression an artifact of recall? J Psychiatr Res 25:141–151

Winokur G, Tsuang M, Crowe RR (1982) The Iowa 500: affective disorder in relatives of manic and depressed patients. Am J Psychiatry 139:209–212

Exhibit 6

Am. J. Hum. Genet. 60:1202–1210, 1997

The Likelihood of Being Affected with Huntington Disease by a Particular Age, for a Specific CAG Size

R. R. Brinkman,[1] M. M. Mezei,[1] J. Theilmann,[1] E. Almqvist,[1,3] and M. R. Hayden[1]

[1]Department of Medical Genetics and [2]Department of Medicine, Division of Neurology, University of British Columbia, Vancouver; and [3]Departments of Molecular Medicine and Clinical Neuroscience, Division of Geriatric Medicine, Karolinska Institute, Stockholm

Summary

Prior studies describing the relationship between CAG size and the age at onset of Huntington disease (HD) have focused on affected persons. To further define the relationship between CAG repeat size and age at onset of HD, we now have analyzed a large cohort of affected and asymptomatic at-risk persons with CAG expansion. This cohort numbered 1,049 persons, including 321 at-risk and 728 affected individuals with a CAG size of 29–121 repeats. Kaplan-Meier analysis has provided curves for determining the likelihood of onset at a given age, for each CAG repeat length in the 39–50 range. The curves were significantly different ($P < .0005$), with relatively narrow 95% confidence intervals (95% CI) (±10%). Penetrance of the mutation for HD also was examined. Although complete penetrance of HD was observed for CAG sizes of ≥42, only a proportion of those with a CAG repeat length of 36–41 showed signs or symptoms of HD within a normal life span. These data provide information concerning the likelihood of being affected, by a specific age, with a particular CAG size, and they may be useful in predictive-testing programs and for the design of clinical trials for persons at increased risk for HD.

Introduction

Huntington disease (HD) is a progressive, neurodegenerative disorder that presents with motor disturbances, psychiatric symptoms, and cognitive decline (Hayden 1981; Harper 1991). Although the mean age at onset is ~40 years of age, 5%–10% of cases have a juvenile onset before age 20 years, whereas late onset (at >50 years of age) occurs in ~20% (Myers et al. 1985). The disease is inexorably progressive, with a duration of ~15–20 years (Hayden 1981; Harper 1991).

Received November 27, 1996; accepted for publication February 21, 1997.

Address for correspondence and reprints: Dr. Michael R. Hayden, Department of Medical Genetics, University of British Columbia, 416-2125 East Mall, NCE Building, Vancouver, B.C. V6T 1Z4 Canada. E-mail: mrh@ulam.generes.ca

0002-9297/97/6005-0022$02.00

The mutation associated with clinical manifestations of the disease is an unstable CAG trinucleotide expansion in exon 1 of a novel gene located on 4p16.3 (Huntington's Disease Collaborative Research Group 1993). Whereas unaffected persons have a range of 6–35 CAG repeats, with 99% of such individuals having <30 repeats, persons affected with HD have CAG sizes of 36–121 repeats (Kremer et al. 1994; Rubinsztein et al. 1996).

In 1994 The Huntington Study Group, a consortium of investigators and practitioners across North America, defined the ranges of CAG repeat length for normal persons as being <30, whereas a CAG of ≥38 indicated a high probability for development of HD (Huntington Study Group 1995). A repeat size of 30–37 was considered "indeterminate" as to whether the patient would develop HD at some time in the future. These guidelines were put into place pending more definitive information concerning the relationship between CAG length and clinical phenotype of the disease. There have been various interpretations of the range of CAG repeat sizes on normal and HD chromosomes (Andrew et al. 1993; Barron et al. 1993; Craufurd and Dodge 1993; De Rooij et al. 1993; Duyao et al. 1993; Goldberg et al.1993*a*; Huntington Disease Collaborative Group 1993; MacMillan et al. 1993; Norremolle et al. 1993; Rubinsztein et al. 1993; Snell et al. 1993; Stine et al. 1993; Zühlke et al. 1993; Benitez et al. 1994; Legius et al. 1994; Trottier et al. 1994; Claes et al. 1995; Soong and Wang 1995), because of different techniques in measuring CAG repeat length, leading to difficulty in comparing results between studies. Originally the DNA analysis consisted of PCR measurement of the CAG repeat, which included the CAG and an adjacent CCG repeat (Goldberg et al. 1993*b*; Huntington's Disease Collaborative Group 1993; Valdes et al. 1993). However, since the CCG-repeat region is polymorphic (Rubinsztein et al. 1993; Andrew et al. 1994), a more accurate PCR technique, which measures only the CAG repeat, has been developed and is the preferred technique for predictive and diagnostic testing (Riess et al. 1993; Warner et al. 1993; Kremer et al. 1995). This has allowed clarification of the ranges of CAG repeat length in persons affected with HD. An analysis of >1,100 affected persons clearly has shown that HD has not occurred with a CAG length <36 repeats (Kremer et al. 1994; Rubinsztein et al. 1996).

However, for patients who are informed that they have inherited the CAG repeat expansion in the HD range (>35), the question often changes from whether they will develop HD to when the disease will manifest. Numerous studies have described a significant inverse relationship between CAG repeat size and age at onset (Andrew et al. 1993; Barron et al. 1993; Craufurd and Dodge 1993; De Rooij et al. 1993; Duyao et al. 1993; Norremolle et al. 1993; Rubinsztein et al. 1993; Simpson et al. 1993; Snell et al. 1993; Stine et al. 1993; Telenius et al. 1993; Zühlke et al. 1993; Ashizawa et al. 1994; Kieburtz et al. 1994; Legius et al. 1994; Novelletto et al. 1994; Trottier et al. 1994; Lucotte et al. 1995; Masuda et al. 1995; Ranen et al. 1995; Yapijakis et al. 1995; Brandt et al. 1996), with the CAG repeat length being invoked to account for ~50% of variation in age at onset (Andrew et al. 1993; Stine et al. 1993; Legius et al. 1994). However, these prior studies of the relationship between age at onset and CAG repeat size have not included asymptomatic, at-risk individuals with CAG length in the affected range. This makes it impossible to take into account the number of asymptomatic persons with a particular CAG at a specific age, and thus it prevents a complete understanding of the relationship between CAG size and age at onset of HD. Both to further explore the relationship between CAG repeat size and the age at onset of clinical manifestations of HD and to determine the risk of developing HD by a certain age in a patient with a particular CAG repeat size, we analyzed CAG size in large numbers of affected and asymptomatic at-risk individuals. Strict criteria were used to determine the age at onset by incorporation of careful clinical follow-up of patients.

Inclusion of the asymptomatic at-risk individuals with expanded CAG repeat lengths, along with use of the more accurate CAG repeat measurement, allowed us to develop estimates of the probability of developing HD, by a particular age, with a given CAG size. The large numbers of individuals in our database also enabled us to clarify the ranges of normal and HD-associated CAG repeat sizes.

HD previously has been considered to be 100% penetrant (Huntington 1872; Hayden 1981; Harper 1991). However, recent data (Nance 1996; Rubinsztein et al. 1996) suggest that, on rare occasions, HD may indeed be not fully penetrant. This data set also has allowed us to study the penetrance of HD at different CAG sizes and to clearly show that penetrance is modified by CAG size.

Subjects and Methods

Subjects

DNA samples from persons with the diagnosis of HD, their at-risk family members, and unrelated controls have been collected in our laboratory since 1984. To date, we have collected information on 4,934 individuals from 1,193 families. A total of 1,593 individuals in our database are affected with HD, and 2,244 are asymptomatic but at risk (i.e., first- or second-degree relatives of an affected individual). Families from Canada and many parts of the world, including families of European, Asian, black South African, Arab, and South American origin, are represented. Samples from all family members of affected individuals were recruited actively, especially asymptomatic and older first- and second-degree relatives of affected individuals, as part of the predictive-testing program at the University of British Columbia clinical genetics clinic.

For the purpose of this study we used those individuals with CAG expansions of the upper allele that were >28, comprising 728 affected and 321 asymptomatic at-risk individuals from 473 families, whose age at onset or oldest age while still asymptomatic could be ascertained. An accurate assessment of the age at onset was performed through both a retrospective review of patient charts and telephone interviews with patients, family members, genetic counselors, and physicians. Age at onset was defined as the first time at which a patient had either neurological or psychiatric symptoms that represented a permanent change from the normal state. The age used for analysis of all asymptomatic individuals was the oldest age when his or her clinical status was last directly confirmed, either at the genetics clinic in Vancouver or by the local, attending physician. Particular attention was paid to confirmation of current age and clinical status of all asymptomatic, at-risk individuals in the HD database who were >65 years of age.

DNA Analysis and Assessment of CAG Repeats

DNA was extracted from leukocytes by standard procedures (Kunkel et al. 1977). The CAG repeat was assessed for all samples by exclusion of the CCG repeat, by use of PCR analysis with primers that flanked either only the CAG repeat (Kremer et al. 1995) or both the CAG repeat and the CCG repeat (Goldberg et al. 1993b), followed by analysis with primers that flank only the CCG repeat (Andrew et al. 1994). Phasing of the CCG repeat was performed by pedigree studies, when necessary. The CAG repeat size was assessed by comparison with sequenced clones of known CAG size.

Data Analysis

For each year of age and CAG size, all individuals, including both those asymptomatic at-risk and those affected, were taken into account to calculate the cumulative probability of having onset of HD by that age, by use of Kaplan-Meier survival analysis (Splus software version 3.1 release 1, 1992; AT&T). Since we had no affected persons with CAG length <36 repeats, individuals with a CAG repeat length >35 were considered to be a cohort at risk, from birth to either neurological or

Table 1

Distribution of Affected Individuals and Asymptomatic Individuals at Risk for HD Who Have CAG Repeat Size >28

Risk Status	Nos. and Ages for CAG Repeat Size of																							
	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51–121	Overall
At risk:																								
No. of individuals	1	8	8	11	18	9	31	12	4	9	13	47	24	31	24	28	13	5	9	4	4	3	5	321
Average current age (years)	69	42	50	38	41	45	49	50	41	36	41	43	41	39	33	32	29	27	28	27	25	26	19	39
Minimum asymptomatic age (years)	69	18	22	11	6	19	7	9	29	10	12	14	21	18	22	13	16	23	18	19	20	23	16	[illegible]
Maximum asymptomatic age (years)	69	79	83	53	80	78	93	85	54	66	79	78	75	57	52	52	40	33	40	39	33	29	23	[illegible]
SD of current age (years)	NA	22	17	12	20	20	23	23	12	18	22	14	14	11	8	8	7	4	8	9	6	3	3	10
Affected:																								
No. of individuals	0	0	0	0	0	0	0	1	4	2	8	64	74	98	92	95	63	58	39	31	26	13	60	728
Average age at onset (years)								65	57	60	61	56	53	48	43	41	37	36	32	31	28	26	20	41
Minimum age at onset (years)								65	53	35	39	35	33	29	28	25	23	19	17	17	13	16	4	4
Maximum age at onset (years)								65	60	84	71	84	75	65	65	58	48	45	45	45	45	34	35	84
SD of age at onset (years)								NA	3	35	11	10	9	7	8	7	6	5	7	6	9	6	8	13
Total no. of individuals	1	8	8	11	18	9	31	13	8	11	21	111	98	129	116	123	76	63	48	35	30	16	65	1,049

psychiatric onset or until death or last contact (censored). The Kaplan-Meier survival curves were compared by use of log-rank statistics.

There were only 32 individuals with 36–38 CAG repeats, and there were a total of 65 individuals with CAG repeats >50. Therefore, these CAG lengths were excluded from the survival analysis, since the small numbers of individuals at each particular CAG size precluded rigorous statistical analysis.

Results

The distribution of affected and asymptomatic at-risk individuals with CAG repeats >28 is shown in table 1. There were no affected individuals with <36 CAG repeats. The ages of asymptomatic persons with a CAG repeat size of 30–35 are shown in table 2. A total of 963 individuals (728 affected) had CAG repeats >35. Of these, 866 individuals (90%) from 445 families had CAG repeat lengths of 39–50 repeats. The linear association between log age at onset and log CAG size was significant ($P < .001$), with an r^2 value of .73.

The cumulative probability of onset, at 5-year intervals, for a given CAG repeat is shown in table 3, with the complete age distribution for each CAG being shown in figure 1. The mean 95% CI for the probability of age at onset with a given CAG repeat by a particular age was narrow (±10%).

As the CAG repeat length increased from 39 to 50, there was a significant increase in the probability of onset ($P < .0005$) by a given age. For example, whereas an individual with 40 CAG repeats had only a 13% probability of having onset by age 45 years, this increased to 32% for someone with 42 CAG repeats, 73% for 44 repeats, and 100% for a person with 46 CAG repeats (table 3 and fig. 1).

There were no individuals with a CAG repeat length >41 who remained asymptomatic at >56 years of age. This indicated that clinical manifestation of the disease was fully penetrant within a normal life span for this CAG repeat range.

There were, however, several individuals with CAG repeats of 36–41 who had not manifested with symptoms of HD within a normal expected life span (table 4). For example, there were two male individuals, 78 and 85 years of age, with 36 CAG repeats, and there was a male 75 years of age with 39 CAG repeats; all of them were asymptomatic. Furthermore, there was a female with 38 CAG repeats and a male with 40 repeats who were not affected until age 84 years, and there was a male with 41 repeats who was not affected until age 75 years (table 4). All these individuals were of Canadian origin, except for the 78-year-old male with 36 repeats, who was of American origin. However, there

Table 2

Age Distribution of Asymptomatic Individuals with CAG Repeat Size of 30–35

Age (years)	No. of Individuals with CAG Repeat Size of						Total
	30	31	32	33	34	35	
0–19	1	0	1	2	1	3	8
20–29	1	1	2	5	2	4	15
30–39	3	0	1	2	1	4	11
40–49	1	4	6	1	0	5	17
50–59	0	2	1	5	2	7	17
60–64	0	0	0	1	2	1	4
65–69	0	0	0	0	0	0	0
71–74	1	0	0	0	0	2	3
75–79	1	0	0	1	1	1	4
80–84	0	1	0	1	0	1	3
85–89	0	0	0	0	0	2	2
90–95	0	0	0	0	0	1	1
Total	8	8	11	18	9	31	85

were no in
>41 repea
A diffe
significant
individual.
CAG repea
r^2 = .96),
3.4 years
39–50 ra
though on
be affected
years for
CAG repea

Table 3

Cumulative Probability of Onset at Different Ages, for a Given CAG Repeat Size

Age of Subject (years)	Cumulative Probability (95% CI) for CAG Repeat Size of			
	39 (n = 21)	40 (n = 111)	41 (n = 98)	42 (n = 129)
30				.02 (.05–.00)
35		.02 (.05–.00)	.02 (.05–.00)	.05[a] (.09–.01)[a]
40	.07[a] (.20–.00)[a]	.08 (.13–.02)	.12 (.18–.05)	.14 (.20–.07)
45		.13 (.19–.05)	.21 (.30–.12)	.32 (.41–.23)
50	.16 (.33–.00)	.21 (.30–.12)	.38 (.48–.26)	.58 (.66–.47)
55		.36 (.46–.25)	.55 (.64–.42)	.81 (.87–.71)
60		.61 (.70–47)	.77 (.85–.65)	.99 (1.00–.91)
65	.36 (.59–.00)	.80 (.88–.67)	.88[a] (.94–.78)[a]	1.00 (NA)
70	.68 (.87–.20)	.90 (.96–.77)	.94 (.98–.85)	
75	.79[b] (.94–.28)[b]	.95 (.99–.82)	.98 (1.00–.88)	
80				
85		1.00[a] (NA)[a]		

Age of Subject (years)	43 (n = 116)	44 (n = 123)	45 (n = 76)	46 (n = 63)
20				.03 (.07–.00)
25		.01 (.02–.00)	.05 (.10–.00)	.06 (.12–.00)
30	.05 (.10–.01)	.04 (.08–.01)	.17 (.25–.08)	.10 (.17–.02)
35	.18 (.26–.11)	.22 (.30–.14)	.37 (.47–.24)	.41 (.52–.27)
40	.39 (.48–.29)	.49 (.58–.39)	.72 (.81–.59)	.86 (.93–.73)
45	.56 (.65–.45)	.73 (.80–.62)	.91 (.96–.80)	1.00 (NA)
50	.87 (.93–.78)	.89 (.94–.80)	1.00[a] (NA)[a]	
55	.93 (.97–.85)	.96 (.99–.88)		
60	.99[a] (1.00–.91)[a]	1.00[a] (NA)[a]		
65	1.00 (NA)			

Age of Subject (years)	47 (n = 48)	48 (n = 35)	49 (n = 30)	50 (n = 16)
15			.07 (.15–.00)	
20	.04 (.10–.00)	.06 (.13–.00)	.30 (.45–.12)	.19[a] (.36–.00)[a]
25	.16 (.25–.04)	.15 (.26–.02)	.41 (.57–.20)	.39 (.59–.09)
30	.36 (.49–.20)	.46 (.61–.26)	.53 (.68–.30)	.73 (.89–.32)
35	.64 (.76–.45)	.78 (.88–.57)	.77 (.89–.53)	1.00 (NA)
40	.89 (.95–.72)	.89 (.96–.69)	.95[a] (.99–.70)[a]	
45	1.00 (NA)	1.00 (NA)	1.00 (NA)	

[a] Values are for 1 year greater (CAG repeat sizes 42, 43, and 49), 1 year less (CAG repeat sizes 39–41 and 50), or 2 years less (CAG repeat sizes 44 and 45) than the stated interval.
[b] Value is for an individual 71 years of age.

were no instances of nonpenetrance with a CAG length >41 repeats.

A difference of a single CAG repeat length has a significant effect on the expected age at onset for an individual. There was a significant linear trend between CAG repeat length and median age at onset ($P < .001$; $r^2 = .96$), with the median age at onset decreasing by 3.4 years (±0.2) for each CAG length increase in the 39–50 range, as shown in table 5. For example, although only 50% of persons with 40 CAG repeats will be affected by age 59 years, this decreases to age 37 years for 45 CAG repeats and to age 27 years for 50 CAG repeats.

To assess the effect of any possible bias introduced by the inclusion of multiple individuals from 445 families, we randomly selected only two individuals and repeated the analysis. There was no significant difference in the results obtained for 39-49 CAG repeats, suggesting that we did not introduce any obvious bias into our database by including two or more individuals from any one family for this CAG range.

Discussion

Using a large cohort ascertained uniformly and analyzed by use of the same accurate measures for CAG

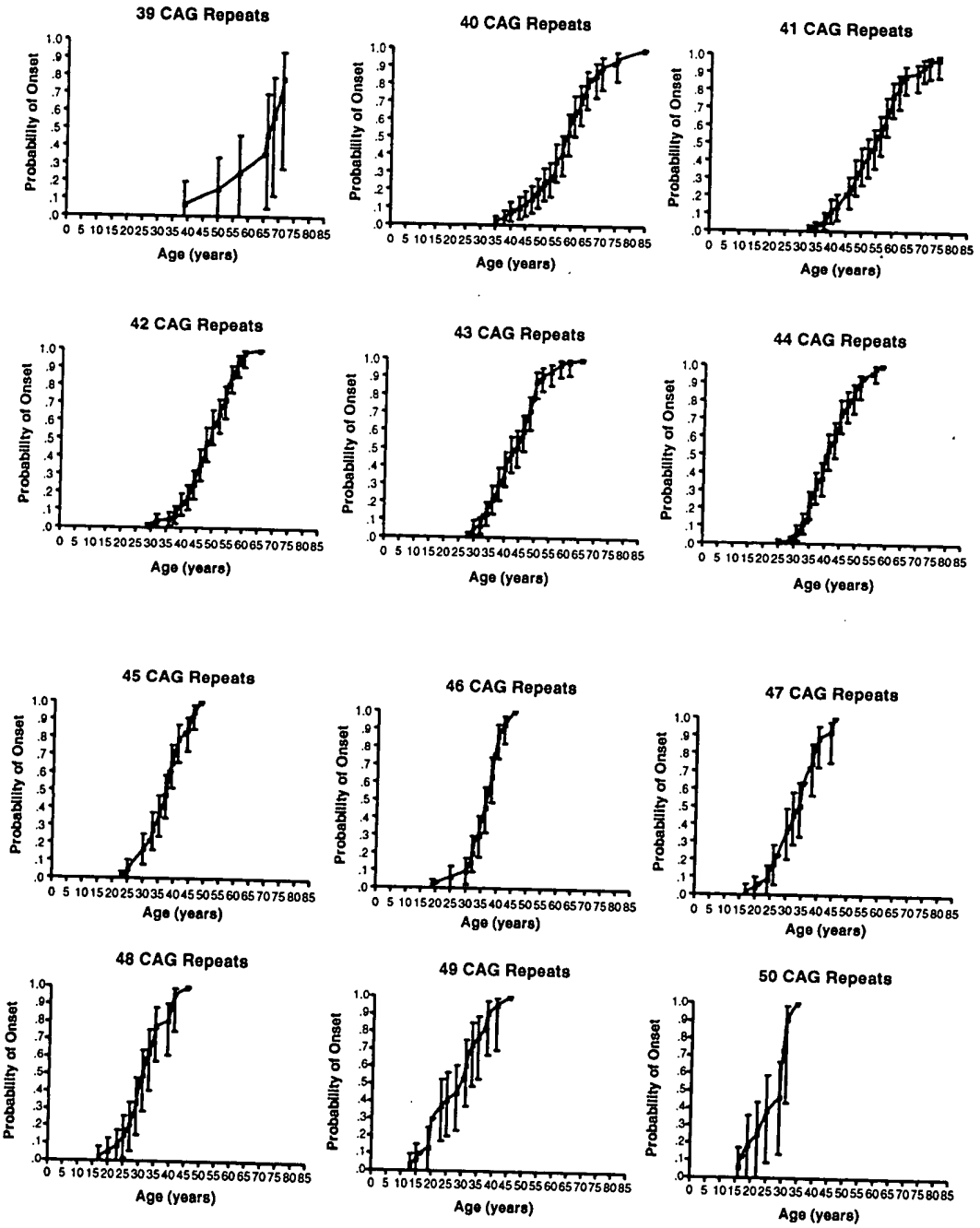


Figure 1 Representations of cumulative probability of being affected, at a particular age, for a CAG of 39–50. Error bars represent 95% CI.

size, this study has clearly shown that CAG repeat length is the major determinant of age at onset in HD. By assessing the CAG size alone, we were able to predict the likelihood that an individual would be affected by a particular age, with relatively narrow 95% CI (±10%) for the vast majority of persons being tested. This study confirms the prior assessments of ranges of CAG repeat length in affected persons and clearly shows that reduced penetrance for HD may occur only for a CAG repeat length <42. The degree of association between CAG size and age at onset was significant ($P < .001$), with an r^2 of .73, which is higher than prior estimates of ~.6 (Andrew et al. 1993; Duyao et al. 1993; Ranen et al. 1995), probably because of definitive determination of CAG size and confirmation of clinical status.

The inverse relationship between the increased CAG repeat size and age at onset for other diseases associated with CAG expansion in novel genes is well documented (La Spada et al. 1992; Jodice et al. 1994; Koide et al. 1994; Nagafuchi et al. 1994; Ranum et al. 1994; Komure et al. 1995; Maciel et al. 1995; Maruyama et al. 1995; Takiyama et al 1995). For HD, since the range

Table 4

Estimation of Penetrance of CAG Expansion in HD Gene, by CAG Repeat Size

CAG Repeat Size	No. of Affected Individuals	No. of Unaffected Individuals, Age >75 Years (Males) or >80 Years (Females)	Total
<29–35	0	9	9
36	1	2	3
37	4	0	4
38	2	1	3[a]
39	8	1	9[a]
40	64	1	65
41	74	1	75
>41	575	0	575

[a] Including individuals reported by Rubinsztein et al. (1996).

of age at onsets for a particular CAG repeat size in the HD gene was very broad, most authors, including ourselves, have recommended against the use of the CAG repeat size to predict the age at onset for an individual patient (Andrew et al. 1993; Barron et al. 1993; Craufurd and Dodge 1993; De Rooij et al. 1993; Duyao et al. 1993; MacMillan et al. 1993; Norremole et al. 1993; Simpson et al. 1993; Stine et al. 1993; Novelletto et al. 1994; Trottier et al. 1994). In this article, including asymptomatic at-risk as well as affected individuals has allowed development of accurate survival curves to predict the probability that an individual will be affected with HD by a certain age. These curves were based on both affected and asymptomatic at-risk individuals with CAG repeat sizes of 39–50, which represents 90% of individuals in our database who have a CAG repeat size in the range of ≥36. The survival curves were significantly different for each CAG repeat size studied, lending further support to the importance of the CAG repeat size as the predominant factor in determining the age at onset. Although these probability curves cannot be used to predict the particular age at onset for an individual, this analysis may have clinical utility by providing estimates of symptom-free survival to an individual seeking additional information in a predictive-testing program.

Data generated from this study also may have significant implications for the design of clinical trials for new therapeutics. The aim of new therapies in HD is to slow or stop the progression of HD in affected persons and to delay or prevent onset in persons with CAG >35 repeats. An appropriate design of clinical trials for increased-risk individuals will need to take into account the expected age at onset of HD for a particular person, to determine the potential efficacy of therapy. For example, all persons with a CAG repeat length of 46 would be expected to manifest with symptoms by 45 years of age, with a median age at onset of 36 years (see table 3 and fig. 1). Extension of age at onset beyond this age could indicate a significant therapeutic effect of a particular drug. Alternatively, shifting of the median age at onset to >36 years also might indicate a therapeutic effect. These data indicate the importance of having sufficient persons with a particular CAG size in a drug trial, allowing for more rapid ascertainment of a beneficial effect.

The results of this study, however, should be interpreted with caution. Although we used a highly accurate methodology to calculate the CAG repeat size, this analysis cannot be extrapolated immediately to other laboratories, because of possible interlaboratory variability in the PCR assays. However, it is somewhat reassuring in this regard that comparison of CAG repeat lengths on the same samples, between laboratories with significant experience in assessment of CAG size, revealed few differences in assessment of CAG repeat size (Marshall and Huntington Study Group 1996; Rubinsztein et al. 1996).

Another cautionary note is that these results were obtained by use of affected and unaffected persons from families with HD. Therefore, these data may not apply equally to asymptomatic individuals in the general population who have no relative with HD but who are found to have CAG repeats in the range seen in affected persons with HD. However, it is indeed extremely rare to find such individuals in the general population. In addition, it should be recognized that a potential bias in the database is underrepresentation of asymptomatic persons at increased risk for developing HD. This would result in underestimation of the age at onset in this study. However, this analysis did include 661 affected and 205 asymptomatic individuals at risk, and, although this does not represent a complete assessment of CAG in a defined population of affected and at-risk individuals, there has been no systematic bias in ascertainment of data.

Table 5

Median Age at Onset

CAG Repeat Size	Median Age at Onset[a] (95% CI) (years)
39	66 (72–59)
40	59 (61–56)
41	54 (56–52)
42	49 (50–48)
43	44 (45–42)
44	42 (43–40)
45	37 (39–36)
46	36 (37–35)
47	33 (35–31)
48	32 (34–30)
49	28 (32–25)
50	27 (30–24)

[a] Age by which 50% of individuals will be affected.

Our results support the recent findings of Rubinsztein et al. (1996), who investigated individuals with 30–40 repeats. We confirm that the lower limit of CAG repeat size in individuals who manifest with HD is 36. This lower limit is supported both by the fact that there were 31 at-risk individuals with a CAG repeat size of 35, including one man who was 93 years old, and that all these persons were asymptomatic (table 2). It is now justified that the CAG size of 30–35 no longer need be considered part of the indeterminate range. These individuals can be informed that there is no clear documentation of any person manifesting with HD who has CAG repeats in this range. However, the risk that offspring will develop HD may be increased, particularly if the transmitting parent is a male (Goldberg et al. 1995; Chong et al. 1997).

Our database includes five males who were asymptomatic at age ≥75 years and one female with onset at 84 years of age (table 4). Penetrance is defined as the proportion of individuals with a specified genotype who show the expected phenotype under a defined set of environmental conditions, which, in this instance, is a normal expected life span (King et al. 1992). In the past, HD was presumed to be 100% penetrant, with all carriers of the HD expansion manifesting the disease (Huntington 1872; Hayden 1981; Harper 1991). The finding of reduced penetrance has been raised previously elsewhere (Nance 1996; Rubinsztein et al. 1996). Using onset at ≥75 years for males and ≥81 years for females as being beyond the normal life span (Statistics Canada 1995), our analysis of survival curves indicates that there is complete penetrance with a CAG repeat size of ≥42. Furthermore, it is apparent that reduced penetrance may occur within the range of 36–41 CAG repeats. Clearly, these data need validation in other independently ascertained large groups of patients, since the numbers are too small to allow for meaningful penetrance estimates for each specific repeat size. However, it is obvious that there is a trend to increasing penetrance with increasing repeat length in the 36–41-repeat range: <90% for 39 CAG repeats and 99% for 41 CAG repeats (table 4).

The data presented in this article provide significant new information for persons and professionals involved in predictive-testing programs. Accurate genotyping provides assessment as to whether a patient has inherited DNA changes associated with HD, but it also can help to define the likelihood of being affected at a certain age in those individuals with a CAG repeat of 39–50. Furthermore, this information also may be useful for improvement of the design of clinical trials of therapeutic agents for increased-risk individuals.

Acknowledgments

The authors are grateful to Drs. Berry Kremer and Jamal Nasir for helpful comments and to Elizabeth Starr for technical assistance. This article also was strengthened by comments from David Oakes of the Huntington Study Group. The authors also wish to acknowledge all doctors and patients who have contributed blood samples to our HD DNA bank and clinical information to the HD database. This work was supported by the Medical Research Council (Canada) and the Huntington Society of Canada.

References

Andrew SE, Goldberg YP, Kremer B, Telenius H, Theilmann J, Adam S, Starr E, et al (1993) The relationship between trinucleotide repeat (CAG) length and clinical features of Huntington disease. Nat Genet 4:398–403

Andrew SE, Goldberg YP, Theilmann J, Zeisler J, Hayden MR (1994) A CCG repeat polymorphism adjacent to the CAG repeat in the Huntington disease gene: implications for diagnostic accuracy and predictive testing. Hum Mol Genet 3: 65–67

Ashizawa T, Wong LJ, Richards CS, Caskey CT, Jankovic J (1994) CAG repeat size and clinical presentation in Huntington's disease. Neurology 44:1137–1143

Barron L, Warner J, Porteous M, Holloway S, Simpson S, Davidson R, Brock DJH (1993) A study of the Huntington's disease associated trinucleotide repeat in the Scottish population. J Med Genet 30:1003–1007

Benitez J, Fernandez E, Garcia Ruiz P, Robledo M, Ramos C, Yebenes J (1994) Trinucleotide (CAG) repeat expansion in chromosomes of Spanish patients with Huntington's disease. Hum Genet 94:563–564

Brandt J, Bylsma FW, Gross R, Stine OC, Ranen NG, Ross CA (1996) Trinucleotide repeat length and clinical progression in Huntington's disease. Neurology 46:527–531

Chong SS, Almqvist E, Telenius H, LaTray L, Nichol K, Bourdelat-Parks B, Goldberg YP, et al (1997) Contribution of DNA sequence and CAG size to mutation frequencies of intermediate alleles for Huntington disease: evidence from single sperm analyses. Hum Mol Genet 6:301–309

Claes S, Van Zand K, Legius E, Dom R, Malfroid M, Baro F, Godderis J, et al (1995) Correlations between triplet repeat expansion and clinical features in Huntington's disease. Arch Neurol 52:749–753

Craufurd D, Dodge A (1993) Mutation size and age at onset in Huntington's disease. J Med Genet 30:1008–1011

De Rooij KE, De Koning Gans PA, Skraastad MI, Belfroid RD, Vegter-van der Vlis M, Roos RA, Bakker E, et al (1993) Dynamic mutation in Dutch Huntington's disease patients: increased paternal repeat instabilities extending to within the normal size range. J Med Genet 30:996–1002

Duyao M, Ambrose C, Myers R, Novelletto A, Persichetti F, Frontali M, Folstein S, et al (1993) Trinucleotide repeat length instability in Huntington disease. Nat Genet 4:387–392

Goldberg YP, Andrew SE, Clarke LA, Hayden MR (1993*a*) A PCR method for accurate assessment of trinucleotide repeat expansion in Huntington disease. Hum Mol Genet 2:635–636

Goldberg YP, Kremer B, Andrew SE, Theilmann J, Graham RK, Squitieri F, Telenius H, et al (1993*b*) Molecular analysis of new mutations causing Huntington disease: intermediate alleles and sex of origin effects. Nat Genet 5:174–17[illegible]

Goldberg YP, McMurray CT, Zeisler J, Almqvist E, Sillence D, Richards F, Gacy AM, et al (1995) Increased instability of intermediate alleles in families with sporadic Huntington disease compared to similar sized intermediate alleles in the general population. Hum Mol Genet 4:1911–1918

Harper PS (1991) Huntington's disease. WB Saunders, London

Hayden MR (1981) Huntington's chorea. Springer, London, Berlin, Heidelberg

Huntington G (1872) On chorea. Med Surg Rep 26:317–321

Huntington Study Group (1995) Huntington Study Group (HSG) policy on non-disclosure of CAG trinucleotide repeat length (CAG_n) genetic (DNA). Huntington Study Group, Rochester, NY

Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72:971–983

Jodice C, Malaspina P, Persichetti F, Novelletto A, Spadaro M, Giunti P, Morocutti C, et al (1994) Effect of trinucleotide repeat length and parental sex on phenotypic variation in spinocerebellar ataxia 1. Am J Hum Genet 54:959–965

Kieburtz K, MacDonald M, Shih C, Feigin A, Steinberg K, Bordwell K, Zimmerman C, et al (1994) Trinucleotide repeat length and progression of illness in Huntington's disease. J Med Genet 872–874

King RA, Rotter JI, Motulsky AG (1992) The genetic basis of common diseases. Oxf Monogr Med Genet 20:3–18

Koide R, Ikeuchi T, Onodera O, Tanaka H, Igarashi S, Endo K, Takahashi H, et al (1994) Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). Nat Genet 6:9–13

Komure O, Sano A, Nishino N, Yamauchi N, Ueno S, Konodoh K, Sano N, et al (1995) DNA analysis in hereditary dentatorubral-pallidoluysian atrophy: correlation between CAG repeat length and phenotypic variation and the molecular basis of anticipation. Neurology 45:143–149

Kremer B, Almqvist E, Theilmann J, Spence N, Telenius H, Goldberg YP, Hayden MR (1995) Sex-dependent mechanisms for expansions and contractions of the CAG repeat on affected Huntington disease chromosomes. Am J Hum Genet 57:343–350

Kremer HPH, Goldberg YP, Andrew SE, Theilmann J, Telenius H, Zeisler J, Squitieri F, et al (1994) A worldwide study of the Huntington's disease mutation: the sensitivity and specificity of measuring CAG repeats. N Engl J Med 330: 1401–1406

Kunkel LM, Smith KD, Boyer SM, Borgaonkar DS, Wachtel SS, Miller OS, Breg WR, et al (1977) Analysis of human Y-chromosome-specific reiterated DNA in chromosome variants. Proc Natl Acad Sci USA 74:1245–1249

La Spada AR, Roling DB, Harding AE, Warner CL, Spiegel R, Hausmanowa-Petrusewicz I, Yee W-C, et al (1992) Meiotic stability and genotype-phenotype correlation of the trinucleotide repeat in X-linked spinal and bulbar muscular atrophy. Nat Genet 2:301–304

Legius E, Cuppens H, Dierick H, Van Zandt K, Dom R, Fryns J-P, Evers-Kiebooms G, et al (1994) Limited expansion of the $(CAG)_n$ repeat of the Huntington gene: a premutation(?). Eur J Hum Genet 2:44–50

Lucotte G, Turpin JC, Riess O, Epplen JT, Siedlaczk I, Loirat F, Hazout S (1995) Confidence intervals for predicted age of onset, given the size of $(CAG)_n$ repeat in Huntington disease. Hum Genet 95:231–232

Maciel P, Gaspar C, DeStefano AL, Silveira I, Coutinho P, Radvany J, Dawson DM, et al (1995) Correlation between CAG repeat length and clinical features in Machado-Joseph disease. Am J Hum Genet 57:54–61

MacMillan JC, Snell RG, Tyler A, Houlihan GD, Fenton I, Cheadle JP, Lazarou LP, et al (1993) Molecular analysis and clinical correlations of the Huntington's disease mutation. Lancet 342:954–958

Marshall FJ, Huntington Study Group (1996) Interlaboratory variability of $(CAG)_n$ determinations in Huntington's disease. Neurology 46:258

Maruyama H, Nakamura S, Matsuyama Z, Sakai T, Doyu M, Sobue G, Seto M, et al (1995) Molecular features of the CAG repeats and clinical manifestation of Machado-Joseph disease. Hum Mol Genet 4:807–812

Masuda N, Goto J, Murayama N, Watanabe M, Kondo I, Kanazawa I (1995) Analysis of triplet repeats in the huntingtin gene in Japanese families affected with Huntington's disease. J Med Genet 32:701–705

Myers RH, Sax DS, Schoenfeld M, Bird ED, Wolf PA, Vonsattel JP, White RF, et al (1985) Late onset of Huntington's disease. J Neurol Neurosurg Psychiatry 48:530–534

Nagafuchi S, Yanagisawa H, Sato K, Shirayama T, Ohsaki E, Bundo M, Takeda T, et al (1994) Dentatorubral and pallidoluysian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. Nat Genet 6:14–18

Nance MA (1996) Huntington disease—another chapter rewritten. Am J Hum Genet 59:1–6

Norremolle A, Riess O, Epplen JT, Fenger K, Hasholt L, Sorenson SA (1993) Trinucleotide repeat elongation in the Huntington gene in Huntington disease patients from 71 Danish families. Hum Mol Genet 2:1475–1476

Novelletto A, Persichetti F, Sabbadini G, Mandich P, Bellone E, Ajmar F, Pergola M, et al (1994) Analysis of the trinucleotide repeat expansion in Italian families affected with Huntington disease. Hum Mol Genet 3:93–98

Ranen NG, Stine OC, Abbott MH, Sherr M, Codori A-M, Franz ML, Chao NI, et al (1995) Anticipation and instability of IT-15 $(CAG)_N$ repeats in parent-offspring pairs with Huntington disease. Am J Hum Genet 57:593–602

Ranum LPW, Chung M-y, Banfi S, Bryer A, Schut LJ, Ramesar R, Duvick LA, et al (1994) Molecular and clinical correlations in spinocerebellar ataxia type 1: evidence for familial effects on the age at onset. Am J Hum Genet 55:244–252

Riess O, Norremoelle A, Soerensen SA, Epplen JT (1993) Improved PCR conditions for the stretch of $(CAG)_n$ repeats causing Huntington's disease. Hum Mol Genet 2:637

Rubinsztein DC, Barton DE, Davison BCC, Ferguson-Smith MA (1993) Analysis of the huntingtin gene reveals a trinucleotide-length polymorphism in the region of the gene that contains two CCG-rich stretches and a correlation between decreased age of onset of Huntington's disease and CAG repeat number. Hum Mol Genet 2:1713–1715

Rubinsztein DC, Leggo J, Coles R, Almqvist E, Biancalana V, Cassiman J-J, Chotai K, et al (1996) Phenotypic characterization of individuals with 30–40 CAG repeats in the Huntington disease gene (HD) reveals HD cases with 36 repeats and apparently normal elderly individuals with 36–39 repeats. Am J Hum Genet 59:16–22

Simpson SA, Davidson MJ, Barron LH (1993) Huntington's disease in Grampian region: correlation of the CAG repeat number and the age of onset of the disease. J Med Genet 30:1014–1017

Snell RG, MacMillan JC, Cheadle JP, Fenton I, Lazarou LP, Davies P, MacDonald ME, et al (1993) Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington's disease. Nat Genet 4:393–397

Soong B-W, Wang JT (1995) A comparison of the Huntington's disease associated trinucleotide repeat between Chinese and white populations. J Med Genet 32:404–408

Statistics Canada (1995) Life tables, Canada and provinces, 1990–1992. Catalogue 84-537. Statistics Canada, Ottawa

Stine OC, Pleasant N, Franz ML, Abbott MH, Folstein SE, Ross CA (1993) Correlation between the onset age of Huntington's disease and length of the trinucleotide repeat in IT-15. Hum Mol Genet 2:1547–1549

Takiyama Y, Igarashi S, Rogaeva EA, Endo K, Rogaev EI, Tanaka H, Sherrington R, et al (1995) Evidence for intergenerational instability in the CAG repeat in the MJD1 gene and for conserved haplotypes at flanking markers amongst Japanese and Caucasian subjects with Machado-Joseph disease. Hum Mol Genet 4:1137–1146

Telenius H, Kremer HPH, Theilmann J, Andrew SE, Almqvist E, Anvret M, Greenberg C, et al (1993) Molecular anal·is of juvenile Huntington disease: the major influence on $(CAG)_n$ repeat length is the sex of the affected parent. Hum Mol Genet 2:1535–1540

Trottier Y, Biancalana V, Mandel JL (1994) Instability of CAG repeats in Huntington's disease: relationship to parental transmission and age of onset. J Med Genet 31:377–382

Valdes JM, Tagle DA, Elmer LW, Collins FS (1993) A simple non-radioactive method for diagnosis of Huntington's disease. Hum Mol Genet 2:633–634

Warner JP, Barron L, Brock DJP (1993) A new poly· . .se chain reaction (PCR) assay for the trinucleotide repeat that is unstable and expanded in Huntington's disease. Mol Cell Probes 7:235–239

Yapijakis C, Vasselopoulos D, Tzagournisakis M, Maris T, Fesdjian C, Papageorgion C, Plaitakis A (1995) Linkage disequilibrium between the expanded $(CAG)_n$ repeat and an allele of the adjacent $(CCG)_n$ repeat in Huntington's disease patients of Greek origin. Eur J Hum Genet 3:228–234

Zühlke C, Riess O, Shroder K, Siedlaczck I, Epplen JT, Engel W, Thies U (1993) Expansion of the $(CAG)_n$ repeat causing Huntington's disease in 352 patients of German origin. Hum Mol Genet 2:1467–1469

 Exhibit 11

Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21

Jean-Louis Blouin[1], Beth A. Dombroski[2], Swapan K. Nath[3], Virginia K. Lasseter[4], Paula S. Wolyniec[4], Gerald Nestadt[4], Mary Thornquist[4], Gail Ullrich[4], John McGrath[4], Laura Kasch[4], Malgorzata Lamacz[4], Marion G. Thomas[4], Corinne Gehrig[1], Uppala Radhakrishna[1], Sarah E. Snyder[2], Katherine G. Balk[2], Karin Neufeld[4], Karen L. Swartz[4], Nicola DeMarchi[5], George N. Papadimitriou[6], Dimitris G. Dikeos[6], Costas N. Stefanis[6], Aravinda Chakravarti[3], Barton Childs[4], David E. Housman[7], Haig H. Kazazian[2], Stylianos E. Antonarakis[1] & Ann E. Pulver[4]

Schizophrenia is a common disorder characterized by psychotic symptoms; diagnostic criteria have been established[1]. Family, twin and adoption studies suggest that both genetic and environmental factors influence susceptibility (heritability is approximately 71%; ref. 2), however, little is known about the aetiology of schizophrenia. Clinical and family studies suggest aetiological heterogeneity[3–6]. Previously, we reported that regions on chromosomes 22, 3 and 8 may be associated with susceptibility to schizophrenia[7–8], and collaborations provided some support for regions on chromosomes 8 and 22 (refs 9–13). We present here a genome-wide scan for schizophrenia susceptibility loci (SSL) using 452 microsatellite markers on 54 multiplex pedigrees. Non-parametric linkage (NPL) analysis provided significant evidence for an SSL on chromosome 13q32 (NPL score = 4.18; *P* = 0.00002), and suggestive evidence for another SSL on chromosome 8p21–22 (NPL = 3.64; *P* = 0.0001). Parametric linkage analysis provided additional support for these SSL. Linkage evidence at chromosome 8 is weaker than that at chromosome 13, so it is more probable that chromosome 8 may be a false positive linkage. Additional putative SSL were noted on chromosomes 14q13 (NPL = 2.57; *P* = 0.005), 7q11 (NPL = 2.50, *P* = 0.007) and 22q11 (NPL = 2.42, *P* = 0.009). Verification of suggestive SSL on chromosomes 13q and 8p was attempted in a follow-up sample of 51 multiplex pedigrees. This analysis confirmed the SSL in 13q14–q33 (NPL = 2.36, *P* = 0.007) and supported the SSL in 8p22–p21 (NPL = 1.95, *P* = 0.023).

A genome scan of 54 families was conducted using 452 markers covering 22 autosomes, a distance of 3,451 centimorgans (cM), with an average resolution of 7.6 cM. We estimate the markers covered more than 90% of the genome when compared with an independent 3,792-cM map (A.C., unpublished data). Using non-parametric analyses, the most significant allele sharing (linkage) between schizophrenics in families is on chromosome 13q32 near marker *D13S174* (NPL = 4.18, *P* = 0.00002), followed by a locus on chromosome 8p21–22 near marker *D8S1771* (NPL = 3.64, *P* = 0.0001; Fig. 1 and Table 1). Additionally, chromosomes 7, 14 and 22 also show NPL scores of approximately 2.5 and *P*-values of less than 0.01. Based on previous criteria[14], the linkage to chromosome 13q is significant, whereas that to chromosome 8p is suggestive.

Parametric analyses were conducted to assess the likelihood of dominant and recessive models. Assuming linkage homogeneity, recessive and dominant models could be excluded for all chromosomes except chromosome 8p, with a peak dominant lod score of 3.19 near *D8S1771*, coincident with the non-parametric evidence. Allowing for locus heterogeneity, on chromosome 8p the peak dominant heterogeneity lod score is 4.54 (70% of families linked); the peak recessive heterogeneity lod score is 2.01 (32% of families linked). On chromosome 13q, the peak dominant heterogeneity lod score is 1.84 (44% of families linked), and the peak recessive heterogeneity lod score is 3.19 (48% of families linked). For chromosomal regions 7q11, 14q13 and 22q11, which have peak NPL scores of 2.5, 2.57 and 2.42 at markers *D7S2212*, *D14S306* and *D22S1265*, respectively, heterogeneity lod scores greater than 1 were also observed. Information content was 80–90% in regions with evidence of linkage.

There are three noted caveats for interpreting the parametric analyses in our study: (i) models tested cannot be true for all segregating susceptibility loci; (ii) if multiple loci are involved, link-

Table 1 • Parametric and non-parametric linkage analysis

	Nonparametric		*Parametric*			
Chromosome Number	NPL	*P*-value	Lod[a] Dom[c]	Lod Rec[d]	HLOD[b] Dom[c]	HLOD Rec[d]
1	1.39	0.084	<–2	<–2	0.18	0.89
2	1.26	0.104	<–2	<–2	0.18	0.48
3	1.59	0.06	<–2	<–2	1.10	0.47
4	0.68	0.247	<–2	<–2	0.03	0.001
5	1.5	0.068	<–2	<–2	0.46	0.75
6	1.96	0.027	<–2	<–2	0.68	0.77
7	2.5	0.007	<–2	<–2	1.30	1.16
8	3.64	0.0001	3.19	<–2	4.54	2.01
9	0.84	0.2	<–2	<–2	0.44	0.09
10	1.87	0.03	<–2	<–2	0.87	1.00
11	1.27	0.102	<–2	<–2	0.43	0.67
12	1.01	0.155	0.12	<–2	1.34	0.50
13	4.18	0.00002	<–2	<–0.40	1.84	3.19
14	2.57	0.005	<–2	<–1.61	0.93	0.44
15	0.9	0.182	<–2	<–2	0.43	0.01
16	0.59	0.276	<–2	<–2	0.06	0.08
17	1.9	0.03	<–2	<–1.84	0.41	0.73
18	0.97	0.166	<–2	<–2	0.16	0.47
19	1.62	0.053	<–2	<–2	0.78	0.41
20	1.22	0.112	<–2	<–2	0.41	0.01
21	1.08	0.139	<–2	<–2	0.34	0.004
22	2.42	0.009	<–2	<–2	1.81	2.10

[a]Lod, lod score; [b]HLOD, heterogeneity lod score; [c]Dom, assumed dominant mode of inheritance; [d]Rec, assumed recessive mode of inheritance.

[1]*Division of Medical Genetics, University of Geneva Medical School and Cantonal Hospital, Geneva, Switzerland, 1211.* [2]*Department of Genetics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.* [3]*Department of Genetics and Center for Human Genetics, Case Western Reserve University and University Hospitals of Cleveland, Cleveland, Ohio 44106, USA.* [4]*Department of Psychiatry and Behavioral Sciences, The Johns Hopkins School of Medicine, Baltimore, Maryland 21231, USA.* [5]*Seconda Universita di Napoli, Instituto di Psichiatria, 80138, Napoli, Italy.* [6]*University Mental Health Research Institute, Eginition Hospital, Athens, Greece, 115 28.* [7]*Center For Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. J.-L.B., B.A.D. & S.K.N. contributed equally to this work. Correspondence should be addressed to A.E.P. (e-mail: aepulver@welchlink.welch.jhu.edu).*

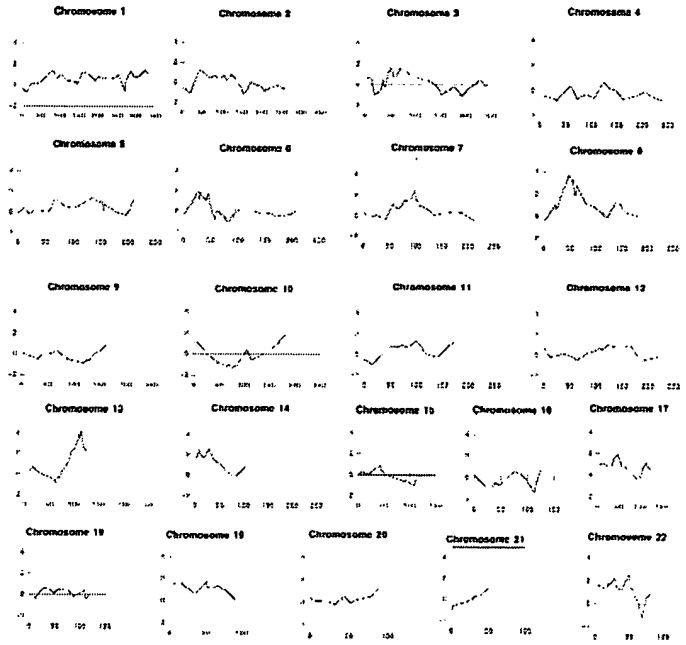


Fig. 1 Genome-wide scan for schizophrenia susceptibility loci (SSLs). Each box contains a figure where the x-axis represents the genetic distance in centimorgans for one of the twenty-two autosomes and the y-axis represents the multipoint non-parametric linkage (NPL) score (Z-all).

age heterogeneity must exist; and (iii) failure to detect linkage is a rejection of the model tested, rather than of linkage itself. First, non-parametric analysis demonstrates significant linkage on chromosome 13 ($P=2.2\times10^{-5}$) and suggestive linkage on chromosome 8 ($P=1.9\times10^{-4}$), but the dominant model (assuming homogeneity) gives a lod score of 3.2 on chromosome 8 and neither model shows significant linkage on chromosome 13. Once linkage heterogeneity is acknowledged, the dominant model has a lod score of 4.5 on chromosome 8 and the recessive model a lod score of 3.2 on chromosome 13. The two susceptibility loci show differing patterns of inheritance. Second, evidence of linkage from parametric analysis only appears upon assuming heterogeneity. The proportions of families demonstrating linkage to chromosome 8 and 13 loci are 70% and 48%, respectively. The total fraction of families showing linkage is 118%, suggesting segregation of multiple factors in the same family. Linkage evidence at chromosome 8 is weaker than that at chromosome 13, so chromosome 8 has a greater probability of being a false positive linkage than chromosome 13. Third, weak linkage on chromosome 22q, by both non-parametric and parametric analyses, suggests that the models chosen were inappropriate for detecting linkage, and a small genetic contribution of this locus in these families.

NPL and lod score analyses on chromosomes 8 and 13 (Fig. 2) show that peak evidence is coincident on each chromosome for both non-parametric and parametric analyses, and that peak location is unchanged when heterogeneity in linkage is assumed. Heterogeneity fits the data (chromosome 8, $\chi^2_1=6.22$, $P<0.02$; chromosome 13, $\chi^2_1=16.53$, $P<0.001$), suggesting that all 54 families do not segregate the putative factors on chromosome 8 and 13. A 95% confidence interval for the region size implicated by parametric linkage analysis is 18 cM for both chromosomes 8 and 13, assuming the dominant and recessive model, respectively. Enlargement of the implicated region for chromosome 8p compared with our previous report[8] may be explained by several factors: (i) a 17-cM interval in the previous map that contained one marker (*D8S136*) has been made more informative by including genotypings of 13 additional polymorphisms, including the locus with the current peak evidence (*D8S1771*); (ii) equal allele frequencies were used for all markers in previous reports, whereas exact allele frequencies are used currently; (iii) 11 families with little linkage information in the previous sample were replaced with eight potentially more informative families; and (iv) multipoint analyses to assess relative pair sharing of alleles increases the amount of linkage information in an interval compared with prior use of two-point analyses.

To add support, we analysed 51 additional families for 17 chromosome 8p22–p21 markers (32 cM) and seven chromosome 13q14–q33 markers (16 cM). For the 8p region, the peak NPL was 1.95 ($P=0.023$) at *D8S1752*, approximately 3 cM telomeric to the *D8S1771* peak in the 54 families. The peak dominant heterogeneity lod score was 0.68 also at *D8S1752*. For the 13q region, the peak NPL was 2.36 ($P=0.007$) at *D13S779*, approximately 7 cM telomeric to the peak at *D13S174* in the 54 families. The peak recessive heterogeneity lod score was 0.61 at *D13S1284*, approximately 5 cM centromeric to the original peak. For both regions, peaks in the NPL curves and in the heterogeneity lod score curves were observed in a flat region extending more than 5 cM.

The difference in the degree of significance between primary and follow-up samples is probably due to random variation in the proportion of linked families at a given SSL. If the true proportion of linked families at a given SSL is 50%, then in a new sample of 50 pedigrees, the proportion of linked families would vary 36–64% with 95% probability.

Our study differs from others in several respects. First, we used a narrow definition of the illness. This may have reduced our statistical power; however, it also may reduce the false positive rate by decreasing the number of statistical tests performed. Second, most families in the primary sample were ascertained from probands in

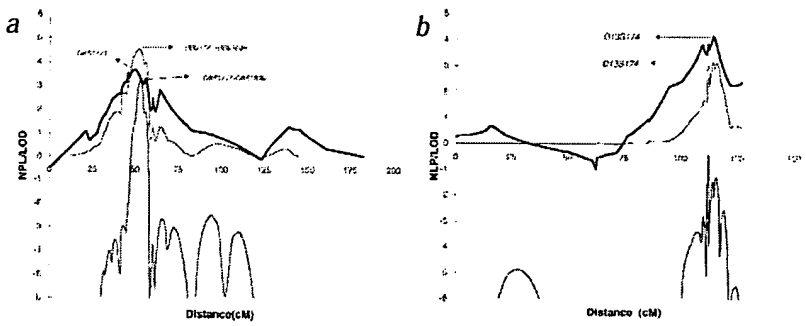


Fig. 2 Multipoint non-parametric and parametric analyses for chromosome 8 (*a*) and chromosome 13 (*b*) from a genome-wide scan for SSLs. The x-axis represents the genetic distance along the chromosome in centimorgans and the y-axis is used to depict the parametric lod score (blue) and the heterogeneity lod score (red), and non-parametric linkage (NPL) score or Z-all (black). For chromosome 8, parametric lod and heterogeneity lod score were obtained using the dominant model. For chromosome 13, parametric lod and heterogeneity lod score were obtained using the recessive model.

Table 2 • Analysis of 10 SZA subjects

Chromosome number	Nonparametric NPL	Nonparametric P-value	Parametric Lod[a]	HLOD[b]
8p: 10 SZA affected	3.50	0.0001	2.16 (Dom[c])	4.14 (Dom)
10 SZA unknown	3.56	0.0001	3.27 (Dom)	4.54 (Dom)
13q: 10 SZA affected	4.18	0.00002	<–0.40 (Rec[d])	3.19 (Rec)
10 SZA unknown	3.25	0.0004	<–2.50 (Rec)	1.93 (Rec)

[a]Lod, lod score; [b]HLOD, heterogeneity lod score; [c]Dom, assumed dominant mode of inheritance; [d]Rec, assumed recessive mode of inheritance.

a systematic sample; affected avuncular pairs were included in addition to affected sibpairs. Third, longitudinal as well as cross-sectional clinical data about each subject from multiple sources was obtained. Longitudinal information helped in differentiation of the psychotic disorders. Fourth, marker maps used loci for which location had high statistical support, because multipoint disease mapping is sensitive to local order and map distances.

In the process of completing the genome scan, linkage reported by others on 6p, 15q and 5q (refs 15–18) was followed up in our sample[19–20]. We failed to detect linkage in these additional regions. Modest evidence for linkage on chromosome 13q14–q32 has been reported[21–22].

To test for linkage heterogeneity, we classified the 54 families in the primary sample according to the expression of psychotic affective disorders or schizophrenic spectrum personality disorders (for example, schizoid, schizotypal and paranoid personality disorders) in their relatives. Six families had 17 relatives with psychotic affective disorders and eight families had 13 relatives with a schizophrenia spectrum personality disorder. Multipoint NPL analyses suggested that families segregating schizophrenia spectrum disorders contributed to linkage findings on both 13q32 (P=0.03) and 8p21 (P=0.00001). There was a negligible impact (P=0.18) of families segregating psychotic affective disorders on the chromosome 8 findings, but this group contributed chromosome 13 findings (P=0.02). The latter data are consistent with evidence for linkage between bipolar illness in multiplex families and markers from chromosome 13q32 (S.D. Detera-Wadleigh, pers. comm.; J. Kelsoe, pers. comm.). Subtyping was not carried out for the follow-up sample. The number of individuals initially seen in the follow-up sample included only those possibly affected and their parents.

To address possible misclassification in the diagnosis of schizoaffective disorder (SZA), we reanalysed the genome scan data on chromosomes 8 and 13, treating the 10 SZA subjects (in nine families) as phenotype unknown. The NPL, P-value, lod score and heterogeneity lod score (Table 2) shows that for chromosome 8, non-parametric evidence remains unchanged; however, the increase in parametric evidence suggests that SZA subjects do not contribute to the chromosome 8 evidence. Treating 10 SZA subjects as unknown in the reanalysis of chromosome 13 reduced both parametric and non-parametric evidence, suggesting SZA classification contributed to the genome scan findings for chromosome 13q. These data are consistent with our heterogeneity subtyping analyses.

To better address the multigenic hypothesis, we computed the correlation coefficient between the NPL scores at the peak locations in the chromosome 8 and 13 SSL for 54 families. The NPL score is independent of any genetic model assumed. In a model of linkage heterogeneity, where some families are linked to the chromosome 8 SSL and others to the chromosome 13 SSL, some families by chance may show segregation at both loci, although the expected correlation must be negative. The observed correlation of 0.12 is positive but not significantly different from zero (P>0.05). We also classified the 54 families based on their NPL scores at chromosome 13 (the SSL with the highest linkage evidence). Thirty-six families had positive chromosome 13 NPL scores (group13+) and 18 families had zero or negative NPL scores (group13–). The peak chromosome 8 NPL score, P-value, lod score and heterogeneity lod score for group13+ was 3.0, 0.0015, 2.3 and 3.2, and for group13– was 2.1, 0.0263, 1.7 and 1.9. Given that group13– is one-half the sample size of group13+ and is expected to have lower information content, these data show that the chromosome 8 SSL evidence arises from both groups. The greater chromosome 8 SSL evidence arises in families likely to be segregating the chromosome 13 SSL. This evidence favors the hypothesis of multigenic inheritance.

The identification of chromosomal regions on 13q32 and 8p21 that are linked to schizophrenia susceptibility is a step towards identifying the genetic aetiology of the disease. The target regions need to be reduced in size by studying additional families, recognizing subtypes that show greater segregation with one or more SSL and saturating the regions with more markers for association and disequilibrium studies.

Methods

Pedigree ascertainment. We identified multiplex families through a systematic sample of schizophrenic patients in the Maryland epidemiologic sample (MES; 44 families of European descent). Details of the MES are available[23,24]. Additional families were identified through nationwide recruitment efforts in the United States (31 families of European descent, 6 African-American, 5 Ashkenazim, 1 Amish) and international collaborations (10 Italian, 3 Polish, 5 Greek). Fifty-seven families were used in our previous report of potential linkage to chromosomes 3 and 8 (ref. 8). The current genome scan sample of 54 families was created by omitting 11 families in which the only affecteds were a parent and a child and adding eight affected sibpair or sib-triplet families.

The genome scan and follow-up samples have a similar number of affecteds per pedigree (2.7 *versus* 2.6) and the same sex distribution among affecteds (three males for every two females). The two samples differed in number of genotyped individuals (363 (80%) *versus* 289 (56%)), age at onset (>35 yr, 15% *versus* 3%), proportion of families with affecteds in more than one generation (50% *versus* 22%) and the inclusion of bilineal families (0% *versus* 4%). These differences may be caused by the majority of families in the follow-up sample being ascertained through recruitment efforts focusing on identifying families with two or more affected siblings. In contrast, the initial genome scan sample was obtained through a systematic sample of families in which the proband and at least one other first or second degree relative was affected.

Clinical assessment methods. An examiner (psychiatrist or psychologist) evaluated the subject's medical and psychiatric history by administration of a semi-structured instrument. Personality characteristics were rated by administration of another semi-structured instrument. All interviews were audiotaped and in addition to the direct interview of each subject, informants were also interviewed. Diagnoses were made according to DSM-IV criteria.

Final diagnoses and age at onset were assigned through a consensus procedure[7]. If there were any disagreement, the two diagnosticians met to resolve any differences. In a small percentage of cases, a third diagnostician was asked to review a case. It was a rare occasion that consensus could not be reached (less than 1% of cases), requiring an individual to be removed from the analysis. We limited our analyses to a narrow and conservative definition of affected (that is, schizophrenia or schizoaffective disorders). Eight percent of the affecteds in 105 families (22 of 276 affecteds) were diagnosed as schizoaffective. Schizoaffectives did not differ from schizophrenics when comparing age at onset, lineality, course or gender. Each multiplex family had at least one schizophrenic and one other first or second degree relative diagnosed with schizophrenia or schizoaffective disorder. Individuals who were given a diagnosis of another psychotic disorder or a schizophrenia spectrum personality disorder were classified as unknown in all analyses.

Markers and maps. In this study we used 452 polymorphic microsatellite markers with average heterozygosity of 76% (range 55–94%). The majority of these markers were dinucleotide repeat polymorphisms, the remainder

being tri- and tetranucleotide repeats. Information regarding the oligonucleotide primers used for genotyping, heterozygosity, allele sizes and frequencies were obtained from databases (http://www.chlc.org, http://www.genethon.fr, http://www.marshmed.org).

No single map that incorporated all loci was available because the genetic markers used arose from multiple sources. Further, accurate map distances between adjacent markers are needed for multipoint linkage analysis due to its sensitivity to false double crossovers arising from genotyping errors[25–26]. Consequently, we reconstructed linkage maps of all human chromosomes using the markers we had chosen[27] based on their published genotype data in the CEPH reference pedigrees (http://www.ceph.fr). The *de novo* human genome genetic linkage map was constructed using MultiMap software[28]. Maps were constructed with odds of support of 1000:1 or greater for markers ranked by heterozygosity. All markers that did not map into a single 1000:1 interval were localized to the most likely interval with 100:1 odds. Further details of the algorithms used for map construction are available[28]. Details concerning the map are available upon request.

Genotypes of polymorphic markers. Genotypes at 452 polymorphic markers were obtained from 363 individuals in the genome scan and at 24 markers from 289 individuals in the follow-up sample. CEPH individuals 1333-01 and 1333-02 were genotyped as controls for allele size. To aid in map construction, eight CEPH families were genotyped for a few markers. For genotyping, all markers were assayed by PCR and polyacrylamide gel electrophoresis. Initially, all alleles were coded within each family. In regions suggestive of linkage on chromosomes 8,13 and 22 with non-parametric *P*-value less than 0.01, marker alleles were recoded across families by allele size. For these latter markers, allele frequencies were estimated from 106 unrelated individuals in the genome-scan sample and from 82 unrelated individuals in the follow-up sample. Laboratory procedures were repeated if necessary to minimize missing genotypes. On average, less than 0.025% of all genotypes for any one marker were missing.

Linkage analysis and genome scanning. We performed parametric and non-parametric multipoint linkage analyses using GENEHUNTER version 1.1 (ref. 29). GENEHUNTER estimates the statistical significance of sharing alleles identical-by-descent between all affected individuals, as well as how much of the total genetic information in a segment has been extracted from the markers studied.

We initially screened the whole genome and performed non-parametric linkage analysis. All marker loci were assumed to have equal allele frequencies, and analysis was based on the map we constructed. Multipoint linkage analysis used GENEHUNTER providing the pointwise *P*-value across each chromosome. These calculations used the non-parametric Z-all statistic, which computes the joint identity-by-descent probability among all affecteds in a family. The *P*-values presented were obtained from GENEHUNTER and were derived by computing the exact probability distribution of the overall Z-all score under the null hypothesis of no linkage. A normal approximation was not used. For regions of chromosomes 8, 13 and 22 with *P*-values less than 0.01, we genotyped additional markers and used the estimated allele frequencies to repeat the analysis. Finally, we also conducted parametric analysis under the assumptions of both homogeneity and linkage heterogeneity. The disease inheritance model used assumed a population prevalence of approximately 1% with approximately 10% phenocopy rate and Hardy-Weinberg equilibrium. For the dominant model this corresponded to penetrances of 0.65, 0.65 and 0.0096 for the mutant homozygote, mutant heterozygote and wild-type homozygote, respectively, and a mutant allele frequency of 0.005; for the recessive model this corresponded to penetrances of 0.65, 0.0096 and 0.0096 for the mutant homozygote, mutant heterozygote and wild-type homozygote, respectively, and a mutant allele frequency of 0.11. Confidence intervals for the size of the regions implicated by linkage were based on the results of parametric analysis and the most likely model using the 1-unit-down method[30]. In this method, all map locations which give a score above Z-1 are considered, Z being the peak lod score.

For both types of analyses, we set the genome-wide false positive rate at 5% and used established criteria[14]. Linkage evidence at a single point in the genome is considered significant whenever the *P*-value is 2.2×10^{-5} or smaller and/or the non-parametric lod score is 3.6 or larger. The evidence is considered suggestive whenever the *P*-value is between 2.2×10^{-5} and 7.4×10^{-4} and/or the non-parametric lod score is 2.2–3.6.

Acknowledgements

This work was supported by funding from Novartis Pharmaceuticals (1995–1998), grants from the National Institutes of Mental Health (NIMH; 1989–1995; R01 MH-45588 and 1 R01 35712), the National Institutes of Health (NIH) grant DRR-OPD-GCRC RR00722, the National Alliance for Research on Schizophrenia and Depression and gifts from private foundations and private donors. We wish to thank the patients and their families and the mental health professionals who referred patients and provided information.

Received 14 April; accepted 5 August, 1998.

1. American Psychiatric Association *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, Washington D.C., 1994).
2. Rao, D.C., Morton, D.E., Gottesman, I.I. & Lew, R. Path analysis of qualitative data on pairs of relatives: Application to schizophrenia. *Hum. Hered.* **31**, 325–333 (1981).
3. Tsuang, M.T. & Faraone, S.V. The case for heterogeneity in the etiology of schizophrenia. *Schizophr. Res.* **17**, 161–175 (1995).
4. Pulver, A.E. & Liang, K.Y. Estimating effects of probands' characteristics on familial risk: II. The association between age at onset and familial risk in the Maryland schizophrenic sample. *Genet. Epidemiol.* **8**, 339–350 (1991).
5. Pulver, A.E. *et al.* Risk factors in schizophrenia: season of birth, gender and familial risk. *Br. J. Psychiatry* **160**, 65–71 (1992).
6. Melton, B., Liang, K.Y. & Pulver, A.E. An extended latent class approach to the study of familial/sporadic forms of disease: Its application to the study of the heterogeneity of schizophrenia. *Genet. Epidemiol.* **11**, 311–327 (1994).
7. Pulver, A.E. *et al.* Sequential strategy to identify a susceptibility gene for schizophrenia: Report of potential linkage on chromosome 22q12-q13.1: Part 1. *Am. J. Med. Genet.* **54**, 36–43 (1994).
8. Pulver, A.E. *et al.* Schizophrenia: A genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes. *Am. J. Med. Genet.* **60**, 252–260 (1995).
9. Pulver, A.E. *et al.* Psychotic illness in patients diagnosed with velo-cardio-facial syndrome and their relatives. *J. Nerv. Ment. Dis.* **182**, 476–478 (1994).
10. Lasseter, V.K. *et al.* Follow-up report of potential linkage for schizophrenia on chromosome 22q: Part 3. *Am. J. Med. Genet.* **60**, 172–173 (1995).
11. Gill, M. *et al.* A combined analysis of *D22S278* marker alleles in affected sib-pairs: Support for a susceptibility locus for schizophrenia at chromosome 22q12. *Am. J. Med Genet.* **67**, 40–45 (1996).
12. Schizophrenia Linkage Collaborative Group for Chromosomes 3, 6 and 8: A Multicenter Study. Additional support for schizophrenia linkage on chromosome 6 and 8. *Am. J. Med. Genet.* **67**, 580–594 (1996).
13. Kendler, K.S. *et al.* Evidence for a schizophrenia vulnerability locus on chromosome 8p in the Irish study of high-density schizophrenia families. *Am. J. Psychiatry* **153**, 1534–1540 (1996).
14. Lander, E. & Kruglyak, L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet.* **11**, 241–247.
15. Straub, R.E. *et al.* A potential vulnerability locus for schizophrenia on chromosome 6p24-p22: Evidence for genetic heterogeneity. *Nature Genet.* **11**, 287–293 (1995).
16. Freedman, R. *et al.* Linkage of a neurophysiological deficit in schizophrenia to a chromosome 15 locus. *Proc. Natl Acad. Sci. USA* **94**, 587–592 (1997).
17. Straub, R.E., MacLean, C.J., O'Neill, F.A., Walsh, D. & Kendler, K.S. Support for a possible schizophrenia vulnerability locus in region 5q22-31 in Irish families. *Mol. Psychiatry* **2**, 148–155 (1997).
18. Wildenauer, D.B. *et al.* Searching for susceptibility genes in schizophrenia by genetic linkage analysis. *Cold Spring Harb. Symp. Quant. Biol.* **61**, 845–850 (1996).
19. Antonarakis, S.E. *et al.* Schizophrenia susceptibility and chromosome 6p24-p22. *Nature Genet.* **11**, 235–236 (1995).
20. Antonarakis, S.E. *et al.* Lack of linkage or association between schizophrenia and the polymorphic trinucleotide repeat within the KCNN3 gene on chromosome 1q21. *Am. J. Med.Genet.* in press.
21. Lin, M.W. *et al.* Suggestive evidence for linkage of schizophrenia to markers on chromosome 13q14.1-q32. *Psychiatr. Genet.* **5**, 117–126 (1995).
22. Lin, M.W. *et al.* Suggestive evidence for linkage of schizophrenia to markers on chromosome 13 in Caucasian but not Oriental populations. *Hum. Genet.* **99**, 417–420 (1997).
23. Pulver, A.E. *et al.* The Johns Hopkins University Schizophrenia Study: An epidemiologic-genetic approach to test the heterogeneity hypothesis and identify schizophrenia susceptibility genes. *Cold Spring Harb. Symp. Quant. Biol.* **61**, 797–814 (1996).
24. Pulver, A.E. & Bale, S.J. Availabilty of schizophranic patients and their families for genetic linkage studies: Findings from the Maryland epidemiology study. *Genet. Epidemiol.* **6**, 671–680 (1989).
25. Lasher, L., Refer, J. & Chakravarti, A. Effects of genotyping error on the estimation of chromosome map length. *Am. J. Hum. Genet.* **49**, 369 (1991).
26. Buetow, K. H. Influence of aberrant observations on high-resolution linkage analysis outcomes. *Am. J. Hum. Genet.* **49**, 985–994 (1991).
27. Dausset, J. *et al.* Centre d'etude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* **6**, 575–577 (1990).
28. Matise, T.C., Perlin, M. & Chakravarti, A. Automated construction of genetic linkage maps using an expert system (Multimap): a human genome linkage map. *Nature Genet.* **6**, 384–390 (1994).
29. Kruglyak, L., Daly, M.J., Reeve-Daly, M.P. & Lander, E.S. Parametric and nonparametric linkage analysis: A unified multipoint approach. *Am. J. Hum. Genet.* **58**, 1347–1363 (1996).
30. Conneally, P.M. *et al.* Report of the committee on methods of linkage analysis and reporting. *Cytogenet. Cell Genet.* **40**, 356–359 (1985).

Exhibit I

Hum Genet (1997) 99:417–420

RAPID COMMUNICATION

Ming-Wei Lin · Pak Sham · Hai-Gwo Hwu
David Collier · Robin Murray · John F. Powell

Suggestive evidence for linkage of schizophrenia to markers on chromosome 13 in Caucasian but not Oriental populations

Received: 13 September 1996

Abstract Previously we reported suggestive evidence for linkage of schizophrenia to markers on chromosome 13q14.1–q32. We have now studied an additional independent sample of 44 pedigrees consisting of 34 Taiwanese, 9 English and 1 Welsh family in an attempt to replicate this finding. Narrow and broad models based on Research Diagnostic Criteria or the Diagnostic and Statistical Manual of Mental Disorders, third edition, revised, were used to define the schizophrenia phenotype. Under a dominant genetic model, two-point lod scores obtained for most of the markers were negative except that marker D13S122 gave a total lod score of 1.06 ($\theta = 0.2$, broad model). As combining pedigrees from different ethnic origins may be inappropriate, we combined this replication sample and our original sample, and then divided the total sample into Caucasian (English and Welsh pedigrees) and Oriental (Taiwanese and Japanese pedigrees) groups. The Caucasian pedigrees produced maximized admixture two-point lod scores (A-lod) of 1.41 for the marker D13S119 ($\theta = 0.2$, $\alpha = 1.0$) and 1.54 for D13S128 ($\theta = 0$, $\alpha = 0.3$) with nearby markers also producing positive A-lod scores. When five-point model-free linkage analysis was applied to the Caucasian sample, a maximum lod score of 2.58 was obtained around the markers D13S122 and D13S128, which are located on chromosome 13q32. The linkage results for the Oriental group were less positive than the Caucasian group. Our results again suggest that there is a potential susceptibility locus for schizophrenia on chromosome 13q14.1–q32, especially in the Caucasian population.

M.-W. Lin (✉) · P. Sham · D. Collier · R. Murray · J. F. Powell
Departments of Psychological Medicine and Neuroscience, Institute of Psychiatry, De Crespigny Park, Denmark Hill, London SE5 8AF, UK
Tel.: +44-171-919-3645; Fax: +44-171-701-9044
e-mail: mlin@hgmp.mrc.ac.uk

H.-G. Hwu
Department of Psychiatry, College of Medicine, National Taiwan University, Taipei, Taiwan, Republic of China

Schizophrenia is a serious mental illness that may be caused by abnormal brain development (Jones and Murray 1991). The lifetime morbid risk worldwide ranges from 0.5 to 1% (Gottesman 1991). The aetiology of schizophrenia is unclear, but family, twin and adoption studies have shown that genetic factors account for approximately 70% of the variance in liability (Gottesman 1991). Nevertheless, the mode of inheritance of the disorder remains unclear. Although recent statistical modelling suggests at least three common genes that act multiplicatively on the risk of illness (Risch 1990), the existence of major genes in some families remains a possibility. Linkage studies have identified several "hot-spots" for schizophrenia: chromosome 22 (Pulver et al. 1994a, b; Lasseter et al. 1995; Vallada et al. 1995; Schizophrenia Collaborative Linkage Group 1996), chromosome 3p (Pulver et al. 1995), chromosome 8p (Pulver et al. 1995) and chromosome 6 (Moises et al. 1995; Schwab et al. 1995; Straub et al. 1995; Wang et al. 1995).

Previously we reported suggestive evidence for linkage of schizophrenia to markers on chromosome 13q14.1–q32 (Lin et al. 1995), and recently two other groups have found suggestive evidence for this region (Antonarakis et al. 1996; Kalsi et al. 1996). In order to follow up this provisional finding, an additional independent sample of 44 pedigrees (9 English, 1 Welsh and 34 Taiwanese) were collected. The British small-to-moderate size families were ascertained by referral of patients with schizophrenia who had at least one known, living, first-degree relative affected with the same illness. The Taiwanese nuclear families were recruited by referral of patients with schizophrenia who had at least one known, living sibling affected with the same disease. These studies were approved by the appropriate hospital ethics committees and informed consent was obtained from all study subjects. Diagnoses are based on Research Diagnostic Criteria (RDC); (Spitzer et al. 1978) for the Welsh and English pedigrees and the Diagnostic and Statistical Manual of Mental Disorders, third edition, revised (DSM-III-R); (American Psychiatric Association 1987) for Taiwanese families. A total of 224 individuals including 93 with

Table 1 Two-point linkage results of admixture *(A-Lod)* and model-free lod *(MFLOD)* scores of schizophrenia for markers on chromosome 13 with the narrow model for replication sample ($n = 44$), Caucasian group ($n = 21$) and Oriental group ($n = 36$)

Marker	Replication sample				Caucasian group				Oriental group			
	A-Lod			MFLOD Lod score	A-Lod			MFLOD Lod score	A-Lod			MFLOD Lod score
	A-lod	θ	α		A-lod	θ	α		A-lod	θ	α	
D13S126	0.58	0	0.45	0.024	0.02	0.3	0.40	0.006	0.79	0	0.55	0.04
HTR2A	0.02	0	0.15	0	0	0	0	0	0	0	0	0.30
D13S119	0.02	0	0.05	0	**1.41**	0.2	1.0	**1.19**	0.14	0	0.20	0
D13S144	0	0	0	0	**1.16**	0	0.35	**1.18**	0	0	0	0
D13S160	0.34	0	0.30	0.72	0.86	0	0.35	0.72	0.11	0	0.15	0.51
D13S121	0.02	0.2	0.20	0.07	0.08	0	0.10	0	0.05	0	0.10	0.24
D13S71	0.01	0.3	0.40	0	0.83	0.2	1.0	0.59	0	0	0	0
D13S122	0.53	0.2	1.0	0.86	**1.19**	0	0.45	**1.08**	0.27	0	0.25	0.48
D13S128	0.03	0.3	0.45	0.22	**1.54**	0	0.30	**1.72**	0.01	0	0.05	0.28
D13S64	0.002	0	0.05	0.17	**1.03**	0.1	0.75	**1.14**	0	0	0	0.18
D13S173	0	0	0	0	0.39	0.3	0.90	0.09	0.12	0	0.15	0

schizophrenia and 17 with other psychiatric disorders were studied. Ten highly polymorphic microsatellite markers and the biallelic HTR2A marker (Warren et al. 1993), which showed positive lod scores from our previous findings, were genotyped using standard polymerase chain reaction (PCR) techniques (Saiki et al. 1985). Genotyping data was first used to construct a genetic map using the program CRI-MAP (Lander and Green 1987). This enabled potential errors appearing as double recombination events in a small genetic distance to be identified and checked. Several rounds of checking and map building were followed until a consistent map was obtained.

Two diagnostic models based on RDC or DSM-III-R were used to define the disease phenotype. A narrow model consisted of RDC or DSM-III-R schizophrenia as affected and a broad model included diagnoses of schizophrenia, schizoaffective disorder and unspecified functional psychosis. Two-point lod score analyses were performed using the program MLINK from the LINKAGE package (Lathrop et al. 1984) and tabulated family by family. As locus heterogeneity was expected, the admixture lod score (A-lod), which assumes that a proportion α of families are linked, was calculated for each marker. Allelic frequencies were calculated from 79 unrelated individuals in the pedigrees. A dominant model was used with a disease gene frequency (q) of 0.008, with penetrance (f_1, f_2) of 0.5 and sporadic risk (f_0) of 0.005 for the narrow model and 0.7 and 0.01 for the broad model. When the narrow model was used, subjects who would have been counted as unaffected under the broad model were designated as phenotype unknown. In this replication sample, two-point total lod scores obtained for most of the markers were negative at a variety of recombination fractions with both diagnostic models except for the marker D13S122, which produced a lod score of 1.06 ($\theta = 0.2$, with the broad model). Because of the uncertainty of the mode of transmission of the disease, the model-free method of linkage analysis, MFLINK (Curtis and Sham 1995), was also applied to our data. Two-point admixture and model-free lod (MFLOD) scores for markers on chromosome 13 with the narrow model for the replication sample are shown in Table 1.

As genes predisposing to a disease may differ between ethnic groups, combining samples from different ethnic origins may be inappropriate. Therefore we divided this replication sample and our original sample into Caucasian (English and Welsh) and Oriental (Taiwanese and Japanese) groups. The 21 Caucasian families used in this group consisted of 171 individuals, including 61 with schizophrenia and 17 with other psychiatric disorders. The 36 Oriental families were composed of 193 individuals, including 85 with schizophrenia and 9 with other psychiatric disorders.

The Caucasian pedigrees gave total two-point lod score of 1.41 and 1.47 for markers D13S119 and D13S128 at $\theta = 0.2$ with the narrow model, respectively. Allowing for heterogeneity, the Caucasian pedigrees produced maximized A-lod scores of 1.41 for marker D13S119 ($\theta = 0.2$, $\alpha = 1.0$), 1.16 for D13S144 ($\theta = 0$, $\alpha = 0.35$), 1.19 for D13S122 ($\theta = 0$, $\alpha = 0.45$), and 1.54 for D13S128 ($\theta = 0$, $\alpha = 0.3$) with nearby markers also producing positive A-lod scores under the narrow model. A two-point model-free lod score of 1.72 was also obtained for D13S128. When we looked at family-by-family linkage results, the majority of the lod score for D13S128 was contributed by two families (lod score = 1.59 for CAR017, lod score = 1.41 for INS035). The results of two-point admixture and model-free lod scores of schizophrenia for markers on chromosome 13 for the Caucasian sample with the narrow model are shown in Table 1.

When five-point model-free linkage analysis was applied to the Caucasian sample, a lod score of 2.58 (narrow model, as shown in Fig. 1) was obtained around markers D13S122 and D13S128, which are located on chromosome 13q32. A small peak of lod score of 1.09 was also observed near markers HTR2A and D13S119. Twelve-point non-parametric linkage (NPL) analysis using the program GENEHUNTER (Kruglyak et al. 1996) was per-

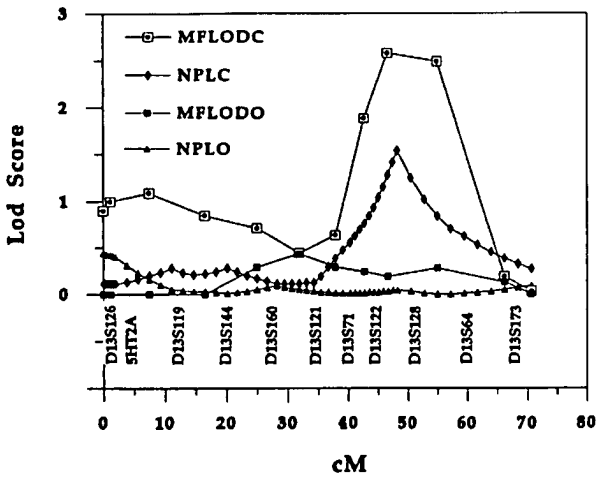


Fig. 1 Multipoint linkage results of schizophrenia for markers on chromosome 13 with narrow model for the Caucasian pedigrees (n = 21) and the Oriental pedigrees (n = 36). *NPL* are 12-point non-parametric linkage analysis using the program GENEHUNTER. *NPLC* indicates lod scores for the Caucasian group and *NPLO* for the Oriental group. *MFLOD* are 5-point model-free lod scores using the program MFLINK coupled with VITESSE. *MFLODC* indicates model-free lod scores for the Caucasian group and *MFLODO* for the Oriental group

formed. For ease of comparison, the NPL statistics and P value produced by GENEHUNTER were transformed to lod score equivalents. A maximum NPL lod score of 1.54 was obtained on the same region as the five-point MFLOD results for the Caucasian group (Fig. 1). The inheritance information extracted from the 11 markers used for the multipoint NPL analysis ranges from 0.74 to 0.91. Although the 12-point non-parametric lod scores obtained were less positive than the 5-point MFLOD scores, the pattern of the NPL results is similar to the MFLOD results.

The linkage results for the Oriental group were less positive than for the Caucasian group. Under the assumption of heterogeneity, a maximum two-point lod score of 0.79 for the marker D13S126 (narrow model) was obtained (Table 1). In the Oriental group, a region around markers D13S160 and D13S121 gave a five-point MFLOD score of 0.43 (Fig. 1). Twelve-point NPL results for the Oriental group shown in Fig. 1 do not suggest linkage.

The lod scores obtained from the replication sample are less impressive than our original findings (an approximate multipoint lod score of 2 was the maximum obtained) with both diagnostic models. However, when the combined sample is divided into different ethnic groups, the results from the Caucasian group, which included ten additional families that are not part of our original sample, are consistent with our previous findings. This was not the case for the Oriental group. We may speculate therefore that the susceptibility locus for schizophrenia in Caucasians may be different from that in Orientals.

The lod score we obtained is below the traditional critical value of 3 for linkage; however, it has been pointed out that some true susceptibility loci may never show significant linkage because they confer a very small increased risk and have common alleles (Owen and Craddock 1996). Two other groups have also found suggestive evidence in this region: Antonarakis et al. (1996) obtained a lod score of 2.54 for D13S128 using a dominant model and Kalsi et al (1996) a lod score of 1.09 for D13S144. It remains possible that there is a dominant susceptibility locus for schizophrenia in a proportion of families, especially for subjects of Caucasian ethnicity. An international collaboration and meta-analysis may be required to elucidate this suggestive linkage evidence to markers on the chromosome 13q14.1–q32 region.

Acknowledgements This work was supported by the Ministry of Education of Taiwan, Republic of China (Ming-Wei Lin), grant DOH (83) (84) (85) (86)-HR-306 from the National Health Research Institute, Taiwan, Republic of China, grant NSC-85-2331-B-002-134 from the National Council of Science, Taiwan, Republic of China (Hai-Gwo Hwu) and Wellcome Trust University Award (Pak Sham). We would like to thank Dr. Mathew Sargeant for the collection of the Welsh pedigree and the HGMP Resource Centre, Hinxton, Cambridge for oligonucleotide synthesis.

References

American Psychiatric Association (1987) Diagnostic and statistical manual of mental disorders: DSM-III-R, 3rd edn, revised. American Psychiatric Association, Washington, DC

Antonarakis SE, Blouin JL, Curran M, Luebbert H, Kazazian HH, Dombroski B, Housman D, et al (1996) Linkage and sib-pair analysis reveal a potential schizophrenia susceptibility gene on chromosome 13q32. Am J Hum Genet 59:A210

Curtis D, Sham PC (1995) Model-free linkage analysis using likelihoods. Am J Hum Genet 57:703–716

Gottesman II (1991) Schizophrenia genesis: the origins of madness. WH Freeman, New York

Jones P, Murray RM (1991) The genetics of schizophrenia is the genetics of neurodevelopment. Br J Psychiatry 158:615–623

Kalsi G, Chen C-H, Smyth C, Brynjolfsson J, Sigmundsson Th, Curtis D, Butler R, et al (1996) Genetic linkage analysis in an Icelandic/British family sample fails to exclude the putative chromosome 13q14.1–q32 schizophrenia susceptibility locus. Am J Hum Genet 59:A388

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

Lander ES, Green P (1987) Construction of multilocus genetic linkage maps in humans. Proc Natl Acad Sci USA 84:2363–2367

Lasseter VK, Pulver AE, Wolyniec PS, Nestadt G, Meyers D, Karayiorgou M, Housman D, et al (1995) Follow-up report of potential linkage for schizophrenia on chromosome 22q: Part 3. Am J Med Genet 60:172–173

Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81:3443–3446

Lin M-W, Curtis D, Williams N, Arranz M, Nanko S, Collier D, McGuffin P, Murray R, Owen M, Gill M, Powell J (1995) Suggestive evidence for linkage of schizophrenia to markers on chromosome 13q14.1–q32. Psychiatr Genet 5:117–126

Moises HW, Yang L, Kristbjarnarson H, Wiese C, Byerley W, Macciardi F, Arolt V, et al (1995) An international two-stage genome-wide search for schizophrenia susceptibility genes. Nat Genet 11:321–324

Owen MJ, Craddock N (1996) Modern molecular genetic approaches to complex traits: implications for psychiatric disorders. Mol Psychiatry 1:21–26

Pulver AE, Karayiorgou M, Wolyniec PS, Lasseter VK, Kasch L, Nestadt G, Antonarakis S, et al (1994a) Sequential strategy to identify a susceptibility gene for schizophrenia: report of potential linkage on chromosome 22q12–q13.1: part 1. Am J Med Genet 54:36–43

Pulver AE, Karayiorgou M, Lasseter VK, Wolyniec P, Kasch L, Antonarakis S, Housman D, et al (1994b) Follow-up of a report of a potential linkage for schizophrenia on chromosome 22q12–q13.1: part 2. Am J Med Genet 54:44–50

Pulver AE, Lasseter VK, Kasch L, Wolyniec P, Nestadt G, Blouin JL, Kimberland M, et al (1995) Schizophrenia: a genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes. Am J Med Genet 60:252–260

Risch N (1990) Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 46:222–228

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anaemia. Science 230:1350–1354

Schizophrenia Collaborative Linkage Group (1996) A combined analysis of D22S278 marker alleles in affected sib-pairs: support for a susceptibility locus for schizophrenia at chromosome 22q12. Am J Med Genet 67:40–45

Schwab SG, Albus M, Hallmayer J, Honig S, Borrmann M, Lichtermann D, Ebstein RP, et al (1995) Evaluation of a susceptibility gene for schizophrenia on chromosome 6p by multipoint affected sib-pair linkage analysis. Nat Genet 11:325–327

Spitzer RL, Endicott J, Robins E (1978) Research diagnostic criteria for a selected group of functional disorders, 3rd edn. New York State Psychiatric Institute, New York

Straub RE, MacLean CJ, O'Neill FA, Burke J, Murphy B, Duke F, Shinkwin R, et al (1995) A potential vulnerability locus for schizophrenia on chromosome 6p24–22: evidence for genetic heterogeneity. Nat Genet 11:287–293

Vallada HP, Gill M, Sham P, Lim LCC, Nanko S, Asherson P, Murray RM, McGuffin P, Owen M, Collier D (1995) Linkage studies on chromosome 22 in familial schizophrenia. Am J Med Genet 60:139–146

Wang S, Sun C, Walczak CA, Ziegle JS, Kipps BR, Goldin LR, Diehl SR (1995) Evidence for a susceptibility locus for schizophrenia on chromosome 6pter–p22. Nat Genet 10:41–46

Warren JT, Peacock JML, Rodriguez LC, et al (1993) An MspI polymorphism in the human serotonin receptor gene (HTR2): detection by DGGE and RFLP analysis. Hum Mol Genet 2:338

Exhibit J

Am. J. Hum. Genet. 65:1096–1103, 1999

Linkage of Familial Schizophrenia to Chromosome 13q32

Linda M. Brzustowicz,[1,2] William G. Honer,[3] Eva W. C. Chow,[4,5] Dawn Little,[1] Jackie Hogan,[5] Kathy Hodgkinson,[5] and Anne S. Bassett[4,5]

[1]Center for Molecular and Behavioral Neuroscience, Rutgers University, and [2]Department of Psychiatry, University of Medicine and Dentistry of New Jersey, New Jersey Medical School, Newark; [3]Department of Psychiatry, University of British Columbia, Vancouver; [4]Department of Psychiatry, University of Toronto, and [5]Genetics Section, Schizophrenia Research Program, Centre for Addiction and Mental Health, Toronto

Summary

Over the past 4 years, a number of investigators have reported findings suggestive of linkage to schizophrenia, with markers on chromosomes 13q32 and 8p21, with one recent study by Blouin et al. reporting significant linkage to these regions. As part of an ongoing genome scan, we evaluated microsatellite markers spanning chromosomes 8 and 13, for linkage to schizophrenia, in 21 extended Canadian families. Families were analyzed under autosomal dominant and recessive models, with broad and narrow definitions of schizophrenia. All models produced positive LOD scores with markers on 13q, with higher scores under the recessive models. The maximum three-point LOD scores were obtained under the recessive-broad model: 3.92 at recombination fraction (θ) .1 with D13S793, under homogeneity, and 4.42 with $\alpha = .65$ and $\theta = 0$ with D13S793, under heterogeneity. Positive LOD scores were also obtained, under all models, for markers on 8p. Although a maximum two-point LOD score of 3.49 was obtained under the dominant-narrow model with D8S136 at $\theta = 0.1$, multipoint analysis with closely flanking markers reduced the maximum LOD score in this region to 2.13. These results provide independent significant evidence of linkage of a schizophrenia-susceptibility locus to markers on 13q32 and support the presence of a second susceptibility locus on 8p21.

Introduction

Schizophrenia is a serious neuropsychiatric illness affecting ~1% of the general population. Family, twin, and adoption studies have demonstrated that schizophrenia is predominantly genetically determined and has high heritability (McGuffin et al. 1994). No gene for schizophrenia has yet been identified, and no significant linkage finding has yet been reproduced in an independent sample (Moldin 1997). Two regions that have recently yielded reports of significant linkage to schizophrenia are chromosomes 13q32 and 8p21 (Blouin et al. 1998). Several groups had previously reported positive, but not significant, findings of linkage to markers on chromosomes 13q (Lin et al. 1995; Pulver et al. 1996; Lin et al. 1997; Straub et al. 1997; Shaw et al. 1998) and 8p (Moises et al. 1995; Pulver et al. 1995; Kendler et al. 1996; Schizophrenia Linkage Collaborative Group for Chromosomes 3, 6, and 8 1996; Kaufmann et al. 1998; Shaw et al. 1998; Wildenauer and Schwab 1998).

Table 1 summarizes positive linkage findings for chromosomes 13q and 8p. The chromosome 13 linkage results cluster within a region of ~12 cM in 13q32, except for the findings of Shaw et al. (1998), which are located within two regions, ~15 cM centromeric on 13q31 (the findings presented in Table 1) and ~50 cM centromeric on 13q12–q13. For all of the chromosome 13 studies, results were obtained by use of narrow definitions of affected phenotype: schizophrenia only or schizophrenia and schizoaffective disorder. Studies that also evaluated evidence for linkage under broader phenotypic models found, in all cases, that a narrowly defined phenotype produced the strongest evidence for linkage to 13q (Lin et al. 1997; Straub et al. 1997; Shaw et al. 1998). Results were not as consistent with respect to recessive and dominant models of inheritance; some studies favored a dominant model, and others a recessive model (table 1). Maximum LOD scores under both inheritance models were not routinely reported, however, making intrastudy comparisons difficult. Interestingly, in the case of two studies using the same set of 54 families but different parametric-linkage methods and marker sets, Pulver et al. (1996) found greater support for linkage to 13q32 under the dominant model of inheritance, whereas Blouin et al. (1998) found greater support under the recessive model of inheritance.

The chromosome 8p linkage results shown in table 1

Received January 15, 1999; accepted for publication July 21, 1999; electronically published September 9, 1999.

Address for correspondence and reprints: Dr. Linda M Brzustowicz, Center for Molecular and Behavorial Neuroscience, Rutgers University, Newark, New Jersey 07102. E-mail: brzustowicz@axon.rutgers.edu

0002-9297/1999/6504-0020$02.00

Table 1

Previous Significant and Suggestive Linkage Findings of Schizophrenia to Chromosomes 13q and 8p

Model-Free Analyses	Parametric Analyses		No. of Families	Reference
	Dominant	Recessive		
Chromosome 13q:				
$P = .00002$[a]	hLOD = 1.84[a]	hLOD = 3.19[a]	54[b]	Blouin et al. (1998)
$P = .0002$[a]	hLOD = 3.24	hLOD = 2.53	54[b]	Pulver et al. (1996)
LOD = 2.58[a]	hLOD = 1.54[a]	...	21	Lin et al. (1997)
$P = .02$[a]	...	hLOD = 1.36[a]	275	Straub et al. (1997)
$P = .03$[a]	hLOD = 1.25	...	70	Shaw et al. (1998)
Chromosome 8p:				
$P = .0001$[a]	hLOD = 4.54[a]	hLOD = 2.01[a]	54[b]	Blouin et al. (1998)
$P = .00004$	LOD = 2.35	LOD = 2.20	57[b]	Pulver et al. (1995)
$P = .006$[a]	hLOD = 2.34[a]	hLOD = 2.52[a]	265	Kendler et al. (1996)
$P = .005$[a]	hLOD = .99	hLOD = 2.22	463[c]	SLCG (1996)
$P = .07$[a]	...	LOD = 1.99	70	Shaw et al. (1998)
$P = .013$[a]	...	...	30	Kaufmann et al. (1998)
$P = .04$	...	...	5	Moises et al. (1995)

NOTE.—P = nominal P values; LOD = maximum homogeneity LOD score; and hLOD = maximum heterogeneity LOD score.

[a] Multipoint analysis result.

[b] The same 54 families were studied by Pulver et al. (1996) and Blouin et al. (1998); 46 of these families overlapped with those studied by Pulver et al. (1995).

[c] No known overlap with samples from other studies referenced.

cluster within a region of ~18 cM, centered around 8p21, except in one study, in which positive results extended over a 50-cM region (Kaufmann et al. 1998). A recent meeting report summarized additional evidence in support of an 8p21 schizophrenia locus (Wildenauer and Schwab 1998), including a LOD score of 2.16, found by Gurling and colleagues, under a dominant model of inheritance, using a set of 23 families from Iceland and the United Kingdom, and an increase in the maximum LOD score to 3.16, in the sample previously reported by Kendler et al. (1996). Most of the chromosome 8p studies reported analyses using a single, narrow definition of schizophrenia (Moises et al. 1995; Pulver et al. 1995; Schizophrenia Linkage Collaborative Group for Chromosomes 3, 6, and 8 1996; Blouin et al. 1998; Kaufmann et al. 1998). Of two studies that considered broader phenotypic definitions, one reported maximum LOD scores under a very broad diagnostic definition, with significant decreases in scores under narrower definitions (Kendler et al. 1996), and another found the strongest evidence for linkage on chromosome 8p using a very narrow definition of disease (Shaw et al. 1998). As with the chromosome 13 findings, positive linkage results on 8p were seen under both dominant and recessive models of inheritance (table 1). Among studies reporting LOD scores for both dominant and recessive models, two studies, using overlapping samples, reported greater support for linkage under the dominant model (Pulver et al. 1995; Blouin et al. 1998). In contrast, two other studies reported higher maximum LOD scores under recessive models of inheritance (Kendler et al. 1996; Schizophrenia Linkage Collaborative Group for Chromosomes 3, 6, and 8 1996).

Despite the differences in the samples used, the range of diagnostic and genetic models tested, and the variety of analytical approaches taken, the multiple findings of positive linkage support the possibility of schizophrenia-susceptibility genes located in relatively small regions of chromosomes 8p and 13q. Given the history of unreplicated findings in the field of psychiatric genetics (Moldin 1997), however, other significant linkage findings in studies using an independent sample of families would help establish these chromosomal regions as more definitively linked to schizophrenia.

As part of an ongoing genome scan, the present study evaluated the evidence for linkage of schizophrenia to chromosomes 8 and 13 in a sample of 21 extended Canadian families. Parametric-linkage analyses were conducted under both autosomal dominant and recessive models, using two diagnostic classifications, narrow and broad. We found significant linkage of schizophrenia to markers on 13q32, providing independent replication of significant linkage results for this chromosomal region.

Subjects and Methods

Subjects

Canadian families of Celtic ($n = 20$) or German ($n = 1$) descent were recruited for the linkage study if schiz-

Table 2

Maximum Pairwise LOD Scores for Schizophrenia and Chromosome 13 Markers

Locus	Marker Position[a]	Hetero-zygosity[b]	Model: Recessive-Narrow LOD (hLOD[c])	Recessive-Narrow θ (α[c])	Recessive-Broad LOD (hLOD)	Recessive-Broad θ (α)	Dominant-Narrow LOD (hLOD)	Dominant-Narrow θ (α)	Dominant-Broad LOD (hLOD)	Dominant-Broad θ (α)
D13S787	0	.80	0 (0)	.4 (.75)	0	.5	.53	.3	.28	.4
D13S1493	17	.75	0 (0)	.5 (.05)	0 (.07)	.4 (.20)	.76	.2	0 (.11)	.5 (.10)
D13S894	24	.68	0	.5	.02 (.03)	.4 (.50)	.09	.4	0	.5
D13S325	30	.79	0	.5	0	.5	0	.5	0	.5
D13S788	37	.85	0	.5	0	.5	0 (.01)	.4 (.65)	0	.5
D13S800	47	.74	.65 (.65)	.2 (.95)	.06	.4	.74 (.96)	.2 (.30)	.21 (.46)	.3 (.25)
D13S317	56	.78	.49	.3	0	.5	1.03 (1.11)	.2 (.80)	0	.5
D13S793	68	.77	2.09 (2.10)	.1 (.95)	1.71 (1.93)	.1 (.55)	1.11 (1.52)	.2 (.40)	.07 (.08)	.4 (.35)
D13S779	75	.65	.77	.2	1.78 (2.20)	.2 (.50)	.82 (1.09)	.2 (.45)	0 (.75)	.5 (.20)
D13S796	86	.81	.31 (.83)	.3 (.25)	1.60 (1.84)	.2 (.50)	.45	.3	.08 (.74)	.4 (.10)
D13S285	103	.80	0	.5	0	.5	0	.5	0	.5

NOTE.—LOD = maximum homogeneity LOD score; hLOD = maximum heterogeneity LOD score.

[a] In Kosambi centimorgans.

[b] Determined from 30 unrelated family members.

[c] For HOMOG analyses with $\alpha < 1$.

ophrenic illness appeared to be segregating in a unilineal autosomal dominant manner (Bassett et al. 1993; Bassett and Honer 1994). Twenty-one moderately large families (n = 285 subjects) were assessed, and 276 subjects had DNA samples available for the current study. All subjects were enrolled in this study after informed consent was obtained, and all procedures conformed to human-subjects protocols approved by the University of Toronto and Rutgers University. The average family size was 22 individuals, with an average of 13 individuals per family participating. Most families consisted of two or three generations with only first- and second-degree relatives participating, although some large complex families contained relatives as distant as ninth degree. Direct interviews—the Structured Clinical Interview for DSM-III-R (SCID-I) for major disorders and the SCID-II for personality disorders— along with collateral information and medical records were used to make consensus diagnoses on the basis of DSM III-R criteria. Further details of the diagnostic and ascertainment methods have been described elsewhere (Bassett et al. 1993; Bassett and Honer 1994). The diagnostic classifications narrow and broad were used, with 71 and 107 affected individuals in each category, respectively. Individuals were considered affected under the narrow diagnostic classification if they were diagnosed with schizophrenia (n = 54) or chronic schizoaffective disorder (n = 17). Individuals were considered affected under the broad diagnostic classification if they had been diagnosed with one of those disorders or with a nonaffective psychotic disorder (n = 13), schizotypal personality disorder (n = 17), or paranoid personality disorder (n = 6). Diagnoses were available on all 285 subjects, and all individuals not categorized as affected under a given diagnostic scheme were classified as unaffected.

Genotyping

DNA was extracted from blood samples or lymphoblastoid cell lines by use of the GenePure system (Gentra Systems). DNA from each subject was genotyped by use of 11 chromosome 13 markers (table 2) and 19 chromosome 8 markers (table 3) from the Weber Screening Set, version 6.0, spanning chromosomes 8 and 13 at an average spacing of 10 cM. Two additional markers from chromosome 8, D8S560 and D8S298, were also genotyped. Genotyping was conducted in our laboratory and the laboratories of the Center for Inherited Disease Research (CIDR; Johns Hopkins University, Baltimore). Approximately 45% of all genotypes were generated by both laboratories. In our laboratory, genotypes were generated by PCR amplification incorporating radiolabeled dCTP. Sample handling, PCR amplification, gel electrophoresis, and genotype-interpretation procedures have all been described elsewhere (Brzustowicz et al. 1997). Genotype generation by CIDR used automated fluorescent microsatellite analysis (further details are available at the CIDR website). The error rate for the CIDR genotypes, calculated from 4,384 genotypes generated for 12 blind duplicate pairs, was 0.08% per genotype. Seventeen genotypes (0.2% of the total used in this study), although concordantly scored on repeat genotyping within and/or between labs, nonetheless were not consistent with Mendelian inheritance. These cases were apparently due to microsatellite instability, but since the source of the original allele could not always

Table 3

Maximum Pairwise LOD Scores for Schizophrenia and Chromosome 8 Markers

Locus	Marker Position[a]	Hetero-zygosity[b]	Model: Recessive-Narrow LOD (hLOD[c])	Recessive-Narrow $\theta(\alpha^c)$	Recessive-Broad LOD (hLOD)	Recessive-Broad θ (α)	Dominant-Narrow LOD (hLOD)	Dominant-Narrow $\theta(\alpha)$	Dominant-Broad LOD (hLOD)	Dominant-Broad $\theta(\alpha)$
D8S264	0	.85	.73 (.74)	.2 (.90)	.26	.3	.35	.3	.30 (.38)	.3 (.70)
D8S1469	15	.82	.03	.4	0	.5	0	.5	.14	.4
D8S1130	21	.71	.57 (.61)	.2 (.80)	.11 (.12)	.3 (.90)	.52 (.53)	.2 (.85)	.32 (.54)	.3 (.35)
D8S1106	25	.69	.72	.2	0	.5	0 (.18)	.5 (.15)	.67 (.88)	.3 (.45)
D8S1145	36	.75	.81 (.81)	.2 (.95)	.39 (.53)	.3 (.30)	.50	.2	.35 (.38)	.3 (.75)
D8S560[d]	43	.80	1.33	.2	.37	.3	.82	.2	.38 (.40)	.2 (.35)
D8S136	43	.80	3.11	.1	.56	.3	3.49	.1	1.29 (1.31)	.2 (.90)
D8S298[d]	44	.72	.90	.2	.98	.2	1.43 (1.47)	.1 (.5)	.87 (1.21)	.3 (.55)
D8S1477	59	.81	.40 (.66)	.3 (.30)	0	.5	.26 (.52)	.3 (.30)	.20 (.20)	.4 (.75)
D8S1110	66	.78	.26 (.27)	.3 (.15)	0	.5	.70	.3	.17	.4
D8S1113	77	.78	.06	.4	.08	.4	0 (0)	.5 (.20)	0	.5
D8S1136	81	.70	.09 (.09)	.3 (.95)	0 (0)	.5 (.50)	.21	.3	0	.5
D8S2324	93	.79	.08	.3	.25	.3	.02	.4	0 (0)	.5 (.25)
D8S1119	100	.83	0	.5	0	.5	.10	.4	.06	.4
C8S14.2	109	.66	.44 (.46)	.2 (.75)	0	.5	0	.5	0	.5
D8S1132	118	.87	.05 (.11)	.3 (.40)	0	.5	.30	.3	0	.5
D8S592	124	.67	.05 (.05)	.4 (.60)	.15 (.21)	.3 (.15)	.15	.3	.08	.4
D8S1179	134	.78	0	.5	0	.5	0	.5	0	.5
D8S1128	138	.75	0	.5	.06	.4	.32	.3	.05 (.05)	.4 (.95)
D8S256	147	.76	.08 (.10)	.3 (.25)	0 (.06)	.5 (.10)	0	.5	0	.5
D8S373	163	.73	.07	.4	.33	.3	0	.5	0	.5

NOTE.—LOD = maximum homogeneity LOD score; hLOD = maximum heterogeneity LOD score.

[a] In Kosambi centimorgans.

[b] Determined from 30 unrelated family members.

[c] For HOMOG analyses with $\alpha < 1$.

[d] Not part of the Weber screening set.

be determined, we consistently deleted these genotypes from our analyses.

Linkage Analysis

Linkage analyses were conducted with the FASTLINK version 4.0P programs (Lathrop and Lalouel 1984; Lathrop et al. 1984, 1986; Cottingham et al. 1993; Schaffer et al. 1994). Parametric-linkage analyses were conducted because they are more powerful than nonparametric methods (Durner et al. 1999) and are robust methods for detecting linkage despite errors or simplifications in the analyzing model, as long as both a dominant and a recessive model are used (Vieland et al. 1992, 1993; Hodge et al. 1997; Greenberg et al. 1998; Durner et al. 1999). To minimize multiple tests, we selected four genetic models to analyze data from the genome scan—one dominant and one recessive for each of the narrow and broad diagnostic classifications. Because of difficulties in determining the appropriate correction for multiple tests when highly correlated diagnostic classifications are used (Lander and Kruglyak 1995), LOD scores are reported with no correction applied. Under the narrow diagnostic classification, the dominant model was schizophrenia susceptibility–allele frequency (p_A) of .0045, with penetrance of disease (f) of .75, .50, and .001 for disease homozygotes (AA), heterozygotes (Aa), and normal homozygotes (aa), respectively; the recessive model was $p_A = .065$, $f(AA) = .50$, $f(Aa) = .0015$, and $f(aa) = .0015$. Under the broad diagnostic classification, the dominant model was $p_A = .007$, $f(AA) = .90$, $f(Aa) = .80$, and $f(aa) = .009$; and the recessive model was $p_A = .10$, $f(AA) = .60$, $f(Aa) = .01$, and $f(aa) = .01$. Marker-allele frequencies for the analysis were estimated with a set of 30 unrelated subjects from these families and were comparable to those listed in the Centre d'Etude du Polymorphisme Humain (CEPH) and Cooperative Human Linkage Center databases. Two-point linkage analyses were conducted by use of MLINK at recombination fraction (θ) 0, .01, .05, .1, .2, .3, and .4. Multipoint analyses were conducted with LINKMAP. Three-point analyses were conducted with adjacent pairs of screening-set markers and the disease locus, with intermarker distances derived from the Marshfield Map positions listed in tables 2 and 3. Additional markers flanking D8S136 were used in four-point analyses with the following map order and recombination fractions: D8S560-.005-D8S136-.0001-D8S298. Heterogeneity testing was conducted with the HOMOG program (Ott 1986, 1991). The large size and complexity of the pedigrees in this sample (Bassett and Honer 1994) made multipoint link-

age analysis by the GENEHUNTER program (Kruglyak et al. 1996) infeasible because of loss of data.

Results

Under all models, two-point linkage analysis produced positive LOD scores with markers on 13q, but with significantly higher LOD scores under the recessive models (table 2). The maximum two-point LOD score of 2.09 under homogeneity was obtained with the recessive model using the narrow diagnostic classification, at $\theta = .10$ with D13S793. The maximum two-point LOD score under heterogeneity was 2.20 at $\theta = .10$ with D13S779 and $\alpha = .5$, obtained under the recessive model using the broad diagnostic classification. A pattern of suggestive positive LOD scores spanning several adjacent markers on chromosome 13q (D13S793, D13S779, and D13S796) was observed (table 2). Inspection of the linkage results by family revealed that the same subset of pedigrees was contributing to the positive linkage findings across this entire region.

Under all models, two-point linkage analysis produced positive LOD scores with markers on 8p but with significantly higher LOD scores under the models using the narrow diagnostic classification (table 3). The maximum two-point LOD score of 3.49 was obtained with the dominant model using the narrow diagnostic classification, at $\theta = .1$ with D8S136, and was the same under homogeneity and heterogeneity. Positive but substantially lower LOD scores were obtained at markers on either side of D8S136 (table 3).

Parametric multipoint analyses of complex disorders must be used with caution, since incorrect analysis models can lead to distortions of the estimated map position of the disease locus and can exclude a true linked locus from the region between close flanking markers (Risch and Guiffra 1992). However, multipoint analyses are useful for combatting the practical limitations caused by inevitable uninformative marker typings, which can either inflate or deflate the LOD score. With large, complex pedigrees, analyzing more than two highly polymorphic marker loci at a time can be computationally prohibitive when large regions of the genome are scanned for linkage. We therefore elected to conduct three-point analyses between adjacent marker loci and the disease locus. For chromosome 13, this produced, across all analysis models under homogeneity, a maximum LOD score of 3.92 at $\theta = .1$ on the centromeric side of D13S793, under the recessive model using the broad diagnostic classification and the marker pair D13S793 and D13S779. In none of the three-point analyses under homogeneity was the maximum LOD score located in the interval between the two marker loci, an anticipated effect of multipoint analysis under an incorrect model. Since one possible error in the model was the assumption of homogeneity, three-point results were also evaluated with HOMOG, for evidence of heterogeneity. Although the hypothesis of heterogeneity was not statistically supported, favored by a likelihood ratio of only 3:1 ($\chi^2 = 2.3$), the location of the maximum LOD score under heterogeneity did shift to the position of one of the marker loci. The maximum three-point LOD score under the assumption of heterogeneity was 4.42, with $\alpha = .65$ at a $\theta = 0$ with D13S793, again under the recessive model using the broad diagnostic classification and the marker pair D13S793 and D13S779 (fig. 1).

For chromosome 8, the maximum three-point LOD score for the screening-set markers, under both homogeneity and heterogeneity, was 3.51 at $\theta = .1$ on the telomeric side of D8S136, under the dominant model using the narrow diagnostic classification and the markers D8S136 and D8S1477. This result was not significantly different from that obtained during the two-point analysis, and the screening-set markers flanking D8S136 (D8S1145 and D8S1477) did not demonstrate supportive positive LOD scores (table 3). Since these findings could indicate a LOD score inflated by uninformativeness of marker D8S136, two closely flanking markers (D8S560 and D8S298) were genotyped. The maximum two-point LOD score under both homogeneity and heterogeneity with D8S560 was 1.33 at $\theta = .2$, under the recessive narrow model (table 3). The maximum two-point LOD score under homogeneity with D8S298 was 1.43 at $\theta = .1$, under the dominant-narrow model, increasing to 1.47 at $\theta = .01$ and $\alpha = .5$, under heterogeneity (table 3). D8S560, D8S136, and D8S298 are very tightly linked, spanning only 0.5 cM and with no evidence of recombination observed in this set of families.

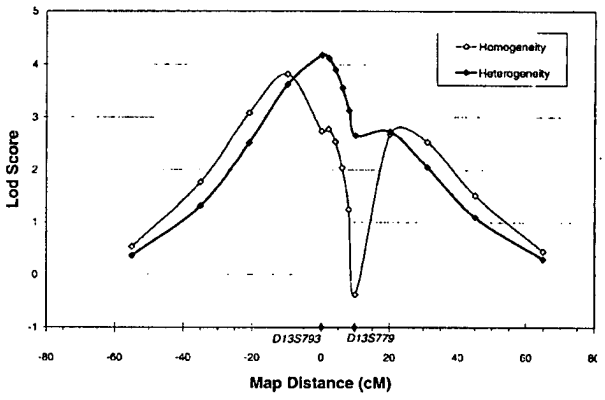


Figure 1 Three-point analysis of schizophrenia, with *D13S793*, and *D13S779* under the recessive-broad genetic model, under homogeneity and heterogeneity. *D13S793* is at map position 0. $\alpha = .65$ for the heterogeneity LOD score.

There was, therefore, no concern that using all three markers in four-point analyses would falsely exclude linkage from the region between these markers. The maximum four-point LOD score under homogeneity across all analysis models was 2.13 at $\theta = .2$ from D8S560, under the recessive model with the narrow diagnostic classification (fig. 2). Results were nearly identical when the dominant-narrow model was used (fig. 2). There was no evidence for heterogeneity in the four-point analyses.

Discussion

To date, individual linkage studies of schizophrenia have produced mixed results, often characterized by suggestive but not significant results and by lack of replication across studies (Moldin 1997). This study produced evidence of a significant linkage on 13q32, as well as of a suggestive linkage on 8p21, <5 cM from the strongest previously reported linkages on both these chromosomes (Blouin et al. 1998). Furthermore, the results independently establish linkage of schizophrenia to markers on chromosome 13q32, satisfying significance standards appropriate for a genomewide scan of a complex disorder.

In this study, we have tested two modes of inheritance and two closely related diagnostic classifications. Lander and Kruglyak (1995) have suggested a LOD score of 3.3 as the significance threshold for linkage studies of complex traits, noting that the appropriate correction for testing multiple, closely related models is not readily known. Morton (1998) has argued that a LOD of 3 provides convincing evidence of linkage even for complex traits, as long as the assumptions of the test are not violated. Hodge et al. (1997) have demonstrated through simulation studies that a 0.3-LOD-unit correction for testing over two dominance models is conservative. Risch's (1991) proposal to subtract $\log_{10}(t)$ for t tests from the obtained LOD score is a conservative correction for analysis over multiple diagnostic or inheritance models and may be overly conservative, given the highly correlated nature of the two affected phenotypes in the current study. Nonetheless, applying a $\log_{10}(t)$ correction for the four tested models to the maximum LOD score of 3.92 for chromosome 13 markers obtained under the hypothesis of homogeneity would still produce a significant corrected LOD score of 3.3.

The autosomal recessive mode of inheritance was favored in our analysis of 13q, as it was in the studies by Straub et al. (1997) and Blouin et al. (1998). This support for an autosomal recessive mode of inheritance was an unexpected result, since one of our ascertainment criteria was a pattern of expression suggestive of autosomal dominant transmission (Bassett and Honer 1994).

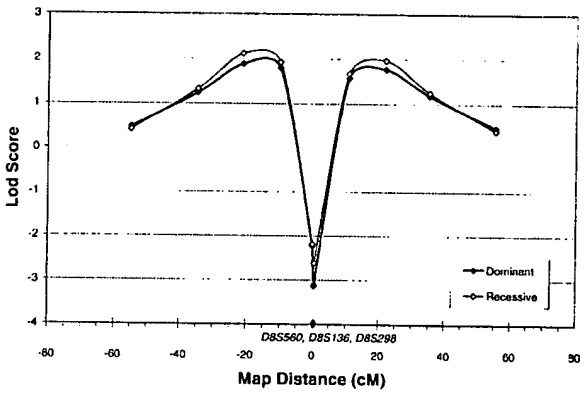


Figure 2 Four-point analysis of schizophrenia, with *D8S560*, *D8S136*, and *D8S298*, under homogeneity and the dominant-narrow and recessive-narrow genetic models. *D8S560* is at map position 0. Because of their close proximity, the positions of the three markers are indicated by a single point, on the *X*-axis, that spans map position 0–0.5 cM. Also because of proximity, the plotted points for D8S136 and D8S298 overlap and appear as a single point on the LOD-score plot.

However, common recessive disease alleles in a population can produce pedigree patterns that appear to be autosomal dominant, underscoring the need for conducting parametric analyses under both dominant and recessive models for disorders with unknown modes of inheritance (Vieland et al. 1992, 1993; Durner et al. 1999).

Although the maximum LOD score for chromosome 13 was obtained by use of the broad definition of affection, prior reports of suggestive and significant chromosome 13 linkages have been obtained by use of a narrow disease definition (Pulver et al. 1996; Lin et al. 1997; Straub et al. 1997; Blouin et al. 1998; Shaw et al. 1998). Although some of these other studies did not perform an analysis under a broad definition of disease (Pulver et al. 1996; Blouin et al. 1998), those that did use multiple diagnostic classifications reported that the largest maximum LOD scores occurred under the narrow definition of disease (Lin et al. 1997; Straub et al. 1997; Shaw et al. 1998). However, since these studies did not also report the scores obtained under the broader diagnostic definitions, it is impossible to gauge the magnitude of the effect of diagnostic classification on maximum LOD score. The maximum homogeneity multipoint LOD score under the recessive narrow model was 3.08, at the same location as the recessive-broad multipoint maximum of 3.92. These scores are significantly higher than the maximum homogeneity multipoint LOD scores obtained under the dominant model: 1.38 and 0.28 with the narrow and broad definitions of disease, respectively.

The maximum LOD scores on 8p were similar under both the dominant- and recessive-narrow models (table 3 and fig. 2), consistent with the results of Pulver et al. (1995) and Kendler et al. (1996). However, there was a striking decrease in LOD scores under the broad models. Most of the prior reports of linkage on 8p considered only a narrow definition of affection (Moises et al. 1995; Pulver et al. 1995; Schizophrenia Linkage Collaborative Group for Chromosomes 3, 6, and 8 1996; Blouin et al. 1998; Kaufmann et al. 1998). Shaw et al. (1998) considered several diagnostic classifications and obtained the strongest evidence for linkage under a very narrow definition of disease. This is opposite to the effect described by Kendler et al. (1996), who reported diminished support for linkage under a narrow diagnostic classification. Better definition of the aspects of the schizophrenia phenotype linked to this locus may help to clarify these results and to define a more appropriate genetic model for future linkage studies.

Recent simulation studies have suggested that, as long as linkage analyses are conducted under both a recessive and dominant model, parametric approaches can have significant power to detect linkage of simple or complex disorders, with little loss of power even when the analyzing model is constructed with an arbitrary disease penetrance (Hodge et al. 1997; Greenberg et al. 1998; Durner et al. 1999). Given this, it is not surprising that despite the slightly different genetic models used by different investigators for their parametric-linkage studies, the same small regions of chromosomes 8 and 13 have been implicated in schizophrenia susceptibility, by a number of groups. It is also clear, however, that not all samples have produced suggestive or significant evidence of linkage to these regions (Coon et al. 1994; Kunugi et al. 1996; Barden and Morissette 1998; Faraone et al. 1998; Levinson et al. 1998; Wildenauer and Schwab 1998), possibly because of factors such as genetic heterogeneity between samples, differences in diagnostic approaches, or sampling effects. The results presented here, as well as the results presented by Blouin et al. (1998), demonstrate that significant linkage results can be obtained in studies of complex disorders when moderate-sized samples (<500 subjects) and parametric-linkage approaches are used. Furthermore, these significant, replicable linkage results suggest that major-gene effects may be present in some samples of familial schizophrenia.

Acknowledgments

We would like to thank the participating families, whose contributions have made these studies possible; Drs. Robert Forsythe and Pamela Forsythe, for years of support; and Manzer Kahn, Jared Hayter, Laura Scutt, and Stephanie Roberts, for technical and administrative assistance. This research was supported by grants from the Medical Research Council of Canada (to A.S.B., L.M.B., and W.G.H.), the EJLB Foundation Scholar Research Programme (to L.M.B.), National Institute of Mental Health (grant K08 MH01392, to L.M.B.), the Ontario Mental Health Foundation (to A.S.B.), the Bill Jeffries Schizophrenia Endowment Fund (to A.S.B.), and the Ian Douglas Bebensee Foundation (to A.S.B.). W.G.H. is supported by a Vancouver Hospital Scientist Award. Genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through contract N01-HG-65403 from the National Institutes of Health to The Johns Hopkins University. Special thanks to Kim Doheny and Elizabeth Pugh of CIDR for their assistance with this project.

Electronic-Database Information

URLs for data in this article are as follows:

Center for Inherited Disease Research, http://www.cidr.jhmi.edu/
CEPH Genotype Database, http://www.cephb.fr/cephdb/
Cooperative Human Linkage Center, http://www.chlc.org/

References

Barden N, Morissette J (1998) Chromosome 13 Workshop. Psychiatr Genet 8:93–96

Bassett AS, Collins EJ, Nuttall SE, Honer WG (1993) Positive and negative symptoms in families with schizophrenia. Schizophr Res 11:9–19

Bassett AS, Honer WG (1994) Evidence for anticipation in schizophrenia. Am J Hum Genet 54:864-870

Blouin JL, Dombroski BA, Nath SK, Lasseter VK, Wolyniec PS, Nestadt G, Thornquist M, et al (1998) Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. Nat Genet 20:70–73

Brzustowicz LM, Honer WG, Chow EWC, Hogan J, Hodgkinson K, Bassett AS (1997) Use of a quantitative trait to map a locus associated with severity of positive symptoms in familial schizophrenia to chromosome 6p. Am J Hum Genet 61:1388–1396

Coon H, Jensen S, Holik J, Hoff M, Myles-Worsley M, Reimherr F, Wender P, et al (1994) Genomic scan for genes predisposing to schizophrenia. Am J Med Genet 54:59–71

Cottingham RW Jr, Idury RM, Schäffer AA (1993) Faster sequential genetic linkage computations. Am J Hum Genet 53: 252–263

Durner M, Vieland VJ, Greenberg DA (1999) Further evidence for increased power of LOD scores over nonparametric methods. Am J Hum Genet 64:281–289

Faraone SV, Matise T, Svrakic D, Pepple J, Malaspina D, Suarez B, Hampe C, et al (1998) Genome scan of European-American schizophrenia pedigrees: results of the NIMH genetics initiative and millenium consortium. Am J Med Genet 81:290–295

Greenberg DA, Abreu P, Hodge SE (1998) The power to detect linkage in complex disease by means of simple LOD-score analyses. Am J Hum Genet 63:870-879

Hodge SE, Abreu PC, Greenberg DA (1997) Magnitude of type I error when single-locus linkage analysis is maximized

over models: a simulation study. Am J Hum Genet 60: 217–227

Kaufmann CA, Suarez B, Malaspina D, Pepple J, Svrakic D, Markel PD, Meyer J, et al (1998) NIMH genetics initiative millenium schizophrenia consortium: linkage analysis of African-American pedigrees. Am J Med Genet 81:282–289

Kendler KS, MacLean CJ, O'Neill A, Burke J, Murphy B, Duke F, Shinkwin R, et al (1996) Evidence for a schizophrenia vulnerability locus on chromosome 8p in the Irish study of high-density schizophrenia families. Am J Psychiatry 153: 1534–1540

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

Kunugi H, Curtis D, Vallada HP, Nanko S, Powell JF, Murray RM, McGuffin P, et al (1996) A linkage study of schizophrenia with DNA markers from chromosome 8p21-p22 in 25 multiplex families. Schizophr Res 22:61–68

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241–247

Lathrop GM, Lalouel J-M (1984) Easy calculations of LOD scores and genetic risks on small computers. Am J Hum Genet 36:460–465

Lathrop GM, Lalouel J-M, Julier C, Ott J (1984) Strategies for multilocus analysis in humans. Proc Natl Acad Sci USA 81:3443–3446

Lathrop GM, Lalouel J-M, White RL (1986) Construction of human genetic linkage maps: likelihood calculations for multilocus analysis. Genet Epidemiol 3:39–52

Levinson DF, Mahtani MM, Nancarrow DJ, Brown DM, Kruglyak L, Kirby A, et al (1998) Genome scan of schizophrenia. Am J Psychiatry 155:741–50

Lin MW, Curtis D, Williams N, Arranz M, Nanko S, Collier D, McGuffin P, et al (1995) Suggestive evidence for linkage of schizophrenia to markers on chromosome 13q14.1-q32. Psychiatr Genet 5:117–126

Lin MW, Sham P, Hwu HG, Collier D, Murray R, Powell JF (1997) Suggestive evidence for linkage of schizophrenia to markers on chromosome 13 in Caucasian but not Oriental populations. Hum Genet 99:417–420

McGuffin P, Asherson P, Owen M, Farmer A (1994) The strength of the genetic effect: is there room for an environmental influence in the aetiology of schizophrenia? Br J Psychiatry 164:593–599

Moises HW, Yang L, Kristbjarnarson H, Wiese C, Byerley W, Macciardi F, Arolt V, et al (1995) An international two-stage genome-wide search for schizophrenia susceptibility genes. Nat Genet 11:321–324

Moldin SO (1997) The maddening hunt for madness genes. Nat Genet 17:127–129

Morton NE (1998) Significance levels in complex inheritance. Am J Hum Genet 62:690–697

Ott J (1986) Linkage probability and its approximate confidence interval under possible heterogeneity. Genet Epidemiol Suppl 1:251–257

——— (1991) Analysis of human genetic linkage, rev. ed. Johns Hopkins University Press, Baltimore

Pulver AE, Lasseter VK, Kasch L, Wolyniec P, Nestadt G, Blouin JL, Kimberland M, et al (1995) Schizophrenia: a genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes. Am J Med Genet 60:252–260

Pulver AE, Wolyniec PS, Housman D, Kazazian HH, Antonarakis G, Nestadt G, Lasseter VK, et al (1996) The Johns Hopkins University Collaborative Schizophrenia Study: an epidemiologic-genetic approach to test the heterogeneity hypothesis and identify schizophrenia susceptibility genes. Cold Spring Harb Symp Quant Biol 61:797–814

Risch N (1991) A note on multiple testing procedures in linkage analysis. Am J Hum Genet 48:1058–1064

Risch N, Giuffra L (1992) Model misspecification and multipoint linkage analysis. Hum Hered 42:77–92

Schaffer AA, Gupta SK, Shriram K, Cottingham RW Jr (1994) Avoiding recomputation in linkage analysis. Hum Hered 44: 225–237

Schizophrenia Linkage Collaborative Group for Chromosomes 3, 6, and 8 (1996) Additional support for schizophrenia linkage on chromosomes 6, and 8: a multicenter study. Am J Med Genet 67:580–594

Shaw SH, Kelly M, Smith AB, Shields G, Hopkins PJ, Loftus J, Laval SH, et al (1998) A genome-wide search for schizophrenia susceptibility genes. Am J Med Genet 81:364–376

Straub RE, MacLean CJ, O'Neill FA, Walsh D, Kendler KS (1997) Genome scan for schizophrenia genes: a detailed progress report in an Irish cohort. Am J Med Genet Neuropsychiatr Genet 74:558

Vieland VJ, Greenberg DA, Hodge SE (1993) Adequacy of single-locus approximations for linkage analyses of oligogenic traits: extension to multigenerational pedigree structures. Hum Hered 43:329–336

Vieland VJ, Hodge SE, Greenberg DA (1992) Adequacy of single-locus approximations for linkage analyses of oligogenic traits. Genet Epidemiol 9:45–59

Wildenauer DB, Schwab SG (1998) Chromosome 8 Workshop. Psychiatr Genet 8:57–87

Exhibit K

The Johns Hopkins University Collaborative Schizophrenia Study: An Epidemiologic-Genetic Approach to Test the Heterogeneity Hypothesis and Identify Schizophrenia Susceptibility Genes

A.E. PULVER,[1] P.S. WOLYNIEC,[1] D. HOUSMAN,[2] H.H. KAZAZIAN,[3] S.E. ANTONARAKIS,[4] G. NESTADT,[1] V.K. LASSETER,[1] J.A. MCGRATH,[1] B. DOMBROSKI,[3] M. KARAYIORGOU,[5] C. TON,[2] J.-L. BLOUIN,[4] AND L. KEMPF[1]

[1]*The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205;* [2]*MIT Center for Cancer Research, Cambridge, Massachusetts 02139;* [3]*University of Pennsylvania Medical Center, Philadelphia, Pennsylvania 19104;* [4]*University of Geneva, Geneva, Switzerland;* [5]*The Rockefeller University, New York, New York 10021*

This paper is divided into two parts. Part I provides an overview of the epidemiology and genetics of schizophrenia and a description of the genetic strategies being pursued to identify schizophrenia susceptibility genes. Part II provides a description of the Johns Hopkins University Collaborative Schizophrenia Study that includes a summary of progress toward the identification and understanding of factors which contribute to schizophrenia susceptibility and to its heterogeneity.

PART I: OVERVIEW OF THE EPIDEMIOLOGY AND GENETICS OF SCHIZOPHRENIA

Epidemiology

Schizophrenia is a frequent and disabling disease of variable expression and unknown cause. Symptoms include hallucinations, delusions, lack of motivation, and anhedonia. Prevalence of the disease is estimated to be about 1% worldwide (Eaton 1985). Ordinarily the illness develops in adolescence or early adulthood but has been observed to have onset in children as well as in later life (Gottesman and Shields 1982). Gender differences can be observed in prevalence, age at onset, premorbid characteristics, clinical expression, course of illness, response to treatment, and family morbidity risk (for review, see Goldstein et al. 1989). In all likelihood, schizophrenia is etiologically heterogeneous (for review, see Tsuang and Faraone 1995).

Risk Factors Associated with Etiology

What risk factors do we know may be associated with cause? Clearly, a family history of schizophrenia increases risk to relatives, and this is the basis for the search for genes involved in susceptibility (Gottesman and Shields 1982). For example, first-degree relatives (parents, siblings, and children) of schizophrenics who share on average 50% of their genes have approximately a 10% risk. Among first-degree relatives, the risk varies from 6% for parents, to 10% for siblings, and 13% for children. If one of the parents is also schizophrenic, the risk to the siblings increases to 17%. Children of two schizophrenics have a risk of 46% of developing the illness. Risk to second-degree relatives is approximately 3%. In addition, it has been repeatedly shown that concordance rates for schizophrenia are higher among monozygotic twins than among dizygotic twins (Gottesman and Shields 1982).

Initially, deficits in attention were shown to predict the later manifestation of schizophrenia among the offspring of schizophrenic parents (Erlenmeyer-Kimling et al. 1993). However, recently, data from a prospective study of a large birth cohort revealed that those individuals who later developed schizophrenia showed deficits in a range of behaviors as early as age four (Done et al. 1994; Jones et al. 1994). This finding implies that childhood and early adolescent behavioral and attentional deficits may be meaningful risk factors for schizophrenia in the general population (i.e., not just for the offspring of schizophrenic parents).

Other factors which may or may not have a genetic basis have been implicated as risk factors. These include season of birth, perinatal or obstetrical complications, and viral infection and/or extreme dietary restriction during the first trimester of pregnancy (for review, see Eaton 1991). With respect to season of birth, there is convincing evidence that a winter–spring birth date increases the risk for schizophrenia (for review, see Boyd et al. 1986). The basis for this association is unknown, but several hypotheses have been offered.

Without knowing the basis for the association between season of birth and the risk for schizophrenia, one can use season of birth to examine the hypothesis that schizophrenia is etiologically heterogeneous by looking for differences between winter-born and non-winter-born schizophrenics which may imply differences in causal mechanisms. Studies using an epidemiological sample of patients hospitalized in the greater Baltimore area have shown that schizophrenic patients who were born during the winter–spring months (February through May) differ from those born in other parts of the year by being more likely to have affected relatives (Pulver et al. 1992b). In addition, mothers of the winter–spring-born schizophrenic

patients were more likely to have a longer interval between their pregnancies than were the mothers of winter–spring-born controls (Pulver et al. 1992a). Other workers have reported that season of birth may be associated with specific clinical features of schizophrenia. For example, Katsanis et al. (1992) provide evidence that winter–spring-born schizophrenics are more prone to dysregulation of electrodermal activity than are winter-born controls. Non-winter-born schizophrenics did not differ from non-winter-born controls with respect to electrodermal activity. In addition, using a large sample of schizophrenic patients, Sacchetti et al. (1992) report that patients born during the winter–spring months show a higher frequency of enlargement of cerebral ventricles when compared to schizophrenic patients born during the rest of the year. The association between season of birth and ventricular enlargement was not present for patients diagnosed with affective disorders.

The role of obstetrical and perinatal events is unclear, but there is overwhelming consistency in the data obtained in retrospective studies that implicate them as risk factors for schizophrenia (for review, see McNeil 1995). More recently, data available from a prospective study of a large birth cohort did not find such an association (Sacker et al. 1995). It remains unclear what specific role these factors play and how they interact with other risk factors such as genetic susceptibility. Susser and colleagues (1996), in a study of a cohort born during the Dutch famine of 1944–45, found that exposure to extreme deficits in nutrition during the first trimester of pregnancy led to an increased risk for schizophrenia in the offspring.

One final association should be mentioned. As reviewed by Eaton et al. (1992), results from several studies suggest a lower than expected rate of rheumatoid arthritis among schizophrenics. The basis for the negative association between the two disorders is unknown. Several theories have been suggested, including nutritional, hormonal, and immunologic.

A Neurodevelopmental Hypothesis

It has been hypothesized that unspecified events, of genetic and/or environmental origin, acting early in fetal life may lead to aberrant development of cerebral structure and function, predisposing to schizophrenic illness later in life (Murray and Lewis 1987; Weinberger 1987; Lewis 1989; Lyon et al. 1989; Murray and O'Callaghan 1991). Although the literature on structural brain abnormalities is not altogether consistent, schizophrenic patients as a group have been reported to have significantly larger ventricles and/or atrophy of the frontal regions when compared to controls (see, e.g., Cannon and Marco 1994). Specific regional changes affecting frontal, temporal limbic, and basal ganglia have been reported (Gur and Pearlson 1993). Neuropathological studies indicate decreased cell numbers and/or cytoarchitectural abnormalities in certain brain regions that suggest disturbed brain-cell migration. The changes do not appear to be associated with degenerative or inflammatory processes and are most likely due to some neurodevelopmental lesion (Benes et al. 1986; Jakob and Beckmann 1986; Weinberger 1987; Jones and Murray 1991; Roberts 1991; Benes 1993; Keshavan et al. 1994; Marsh et al. 1994; Murray 1994; Ross and Pearlson 1996).

Additional evidence to support the neurodevelopmental hypothesis is that, in some cases, enlarged cerebral ventricles are present at or before the onset of the illness, whereas in others there is reduced volume of both temporal lobe and hippocampus (Weinberger 1987). Recent studies reported by Petty et al. (1995) have implicated the heteromodal cortex. Reversed asymmetry of the planum temporale was noted in a higher proportion of the schizophrenic subjects when compared to matched controls. In addition, both dermatoglyphic abnormalities and minor physical anomalies have been found at a higher rate among schizophrenic subjects than normal controls (Green et al. 1994). These findings provide additional support that differences between schizophrenics and the general population are present in utero.

A consensus is thus emerging that this illness occurs because of disturbances in intrauterine life that lead to structural and functional brain abnormalities. These could take the form of abnormalities of cell migration, failure to make connections, or reduced density of synapses, and clinical heterogeneity might vary according to the localization of the abnormality. There is reason to believe that the genetics of schizophrenia (or some forms of schizophrenia) is tied to the genetics of neurodevelopment, an observation that should help to focus efforts to pursue the elusive genetic basis of this devastating disease.

In summary, our understanding of the epidemiology of schizophrenia is far from complete. Additional studies with large numbers of subjects are needed to determine the interrelationships among risk factors and clinically descriptive variables. Most of the studies have emphasized one or two variables, perhaps because of limited numbers in study samples. Samples of sufficient size are necessary to resolve the remaining questions. The samples are only likely to be assembled through collaboration of several investigators. We have not been accustomed to such large-scale collaborative work; if we are to explore the heterogeneity of the disease by associations of constellations of genes with clusters of clinical variables and other risk factors, the assembly of large samples of patients and families is necessary.

Genetics

Twin, adoption, and family studies implicate genetic factors in the etiology of schizophrenia (for reviews, see Gottesman and Shields 1982; Murray et al. 1986; Tsuang et al. 1991). The mode of inheritance is not

known (Risch and Baron 1984; Kendler and Diehl 1993). In all likelihood, inheritance of susceptibility is complex, involving both genes and experiences (see, e.g., McGue and Gottesman 1991). Schizophrenia shows familial co-aggregation with schizoaffective disorder, as well as with atypical psychotic disorders, schizotypal, avoidant, and paranoid personality disorders (Kendler et al. 1984, 1993). In some families, there is co-aggregation of both schizophrenia and bipolar disorder, which raises the possibility of some shared vulnerability (Taylor 1992).

Genetic Strategies to Identify Schizophrenia Susceptibility Genes

Linkage studies. Linkage studies test the association of a specific marker with disease status in a family. Linkage studies of schizophrenia date to as early as 1973 when Elston tested for linkage between schizophrenia and alleles of the Gc and Gm loci (Elston et al. 1973). Only within the past 6 years has a genome-wide search become possible due to the advances in recombinant DNA and other techniques that have led to the availability of an increasing number of highly polymorphic markers located throughout the genome. Although these markers have made it technically possible to complete the molecular analysis, other problems remain in studying complex diseases such as schizophrenia. The lack of sufficiently large numbers of well-characterized patients and their relatives, the need for more complex statistical analysis of the data, the requirement for new computer technology to allow efficient application of analyses to multiple variables, and the continuing need for even more markers in certain regions still limit the success of this work.

Two linkage paradigms have been used: the genome-wide scan and a candidate gene or region approach. (1) *Genome scan:* The genome-wide scan, which provided an early success in identifying the gene responsible for Huntington's disease (Gusella et al. 1983) is labor-intensive and costly. Results from several genome-wide scans have been reported (Barr et al. 1994; Coon et al. 1994b; Pulver et al. 1994c; Moises et al. 1995b). Despite the effort involved, none of these studies had enough families to establish conclusive evidence either for linkage and subsequent pursuit of more extensive cloning strategies or for rejecting linkage under heterogeneity. However, several regions of interest were identified. Cumulative evidence from different populations is now available to suggest that schizophrenia susceptibility genes may be found in regions on chromosome 22q (Coon et al. 1994a; Lasseter et al. 1995), chromosome 8p (Pulver et al. 1995), and chromosome 6p (Straub et al. 1995). Additional results from our group's genome-wide scan are reported in Part II of this paper. (2) *Candidate genes*: In schizophrenia, it is difficult to determine what genes should be considered candidates, given that the pathophysiology of schizophrenia is not known. Many hypothesized "candidate" genes have been tested in linkage studies. A list of hypothesized candidate genes that have been used in linkage and/or association studies is provided in Table 1. No consistent evidence has developed from this linkage approach. However, no candidate gene can be excluded and determined not to have a role in susceptibility to schizophrenia.

A *candidate region* approach may be more productive for schizophrenia. For example, structural chromosomal changes in affected individuals, including translocations and deletions, may lead to the identification of a specific region where a susceptibility gene may lie. This approach proved rewarding in neurofibromatosis, where reports of translocation in two affected individuals helped to localize the responsible gene (Fountain et al. 1989). There are several reports of structural chromosomal changes in schizophrenic subjects. A review of the literature was prepared by Ann Bassett (1992). Balanced and unbalanced translocations, inverted segments, deleted segments, and fragile sites were reported to occur in a single patient or several patients. They were found throughout the genome (i.e., 1q32.3 1q43, 2p11-2q13, 2q21, 2p21, 5p13-[5pter], 5q11.2-5q13.3, 6q14.2, 6q15, 7p12-[7pter], 8p23, 9p11-13, 9p22 10p12-10q21,11q25 [11qter] 13q21.2 13q31-13q32 17p12 18p 18q23 [18qter] 19p13,21) (Bassett 1992). Since her review, additional translocations in schizophrenic patients have been reported: 1p22 (Gordon et al. 1994), 2p11.29 (Maziade et al. 1993), 5p14.1 (Malaspina

Table 1. Hypothesized Candidate Genes for Schizophrenia Examined in Linkage or Association Studies

Candidate	Location
β-Amyloid precursor protein (APP)	21q21.3-q22.05
Catechol *O*-methyltransferase (COMT)	22q11
Debrisoquine hydroxylase (CYP2D6)	22q13.1
Dopamine receptor 1 (DRD1)	5q35.1
Dopamine receptor 2 (DRD2)	11q22-23
Dopamine receptor 3 (DRD3)	3q13.3
Dopamine receptor 4 (DRD4)	11p15.5
Dopamine receptor 5 (DRD5)	4p 15.1-16.1
Dopamine transporter (DAT)	5p15.3
$GABA_A$ receptor B	14p12-13
Glutamate receptor (GLUR5)	21q22.1
Neuropeptide Y (NPY)	7p15.1
Neurotrophin-3 (NT-3)	12p13
Porphobilogen deaminase (PBGD)	11q23.2-qter
Proopiomelanocortin (POMC)	2p25
Serotonin receptor 1A (HTR1A)	5q11.2-q13
Serotonin receptor 2A (HTR2A)	13q14-q21
Serotonin receptor 2C (HTR2C)	Xq24
Serotonin receptor 1D (HTR1DB)	1p36.3-p34.3
Tyrosine hydrogenase (TH)	11p15.5
Vesicular monoamine transporter (VMAT2)	10q25
Dopamine-β-hydroxylase (DBH)	9q34
Interleukin 2 (IL-2)	4q26-q27
Monoamine oxidase (MAO)	Xp11.23
Marfan	15q21.1
Human leukocyte antigens (HLA)(DQB1)	6p21.3

et al. 1992), 11q25 (Holland and Gosden 1990), an inversion on 4p15.2-q21.3 (Palmour et al. 1994), a trisomy of chromosome 5 (Rios et al. 1995), fragile sites at 8q24 10q24 (Garofalo et al. 1993), pericentric inversion of 9p11-9q13 (Nanko et al. 1993), and a deleted segment of 22q11.21-q11.23 (Bland 1982; Cannon-Spoor et al. 1982; Pulver et al. 1994b; Lindsay et al. 1996).

In addition to the candidate regions described above, there is also an hypothesis that the susceptibility to schizophrenia may be related to mutations specific to the sex chromosomes (for review, see Delisi et al. 1994a). Linkage/sib pair studies examining markers in the pseudo-autosomal region of the X chromosome have been reported (see, e.g., Collinge et al. 1991; Asherson et al. 1992; D'Amato et al. 1992, 1994; Wang et al. 1993; Barr et al. 1994; Crow et al. 1994; Delisi et al. 1994b; Kalsi et al. 1995; Maier et al. 1995; Okoro et al. 1995). The evidence from these studies is not conclusive; one cannot rule out the possibility of linkage of some region of the X chromosome with susceptibility to schizophrenia in some families.

Using a linkage strategy to identify schizophrenia susceptibility genes, one cannot expect to be able to identify a specific locus; rather, these studies may identify a region of interest within 1 cM of the gene (see, e.g., Devlin and Risch 1995). Using a combined linkage/association strategy (the association study strategy is described below) is likely to have more success (Ghosh 1995). The application of tests for linkage is not straightforward, given some of the properties of this disease. Difficulties include:

1. *Late age at onset:* Individuals may have the susceptibility gene but not exhibit the disease because they have not lived through the risk period.
2. *Variability in expression:* Uncertainty remains about the definition of a case of schizophrenia. Several diagnostic systems can be employed (see, e.g., Feighner et al. 1972; Endicott et al. 1975; American Psychiatric Association 1994), each with somewhat different criteria for schizophrenia. Pooling samples may cause confusion because of these diagnostic differences. Inter-rater reliability, even when using the same diagnostic system, cannot be assumed.
3. *Incomplete penetrance* has been assumed because concordance rates among monozygotic twins has been reported to vary between 40% and 60%. In addition, a study of the offspring of discordant monozygotic twins suggests that the rate of schizophrenia did not differ when the offspring of affected and nonaffected co-twins were compared, providing support for the hypothesis of incomplete penetrance (Kringlen and Cramer 1989). Using an "affecteds only" design where unaffected individuals have little impact on the analysis can circumvent this problem.
4. *Assortative mating* (i.e., the tendency for mated pairs to be more similar for some phenotypic trait than would be expected if the choice of a partner occurred at random) has been demonstrated to be present among schizophrenic patients (Rao et al. 1981; Parnas 1985, 1988). Nonrandom mating may result in a distortion of the distribution of genotypes in the offspring, and within family heterogeneity.
5. *Reduced fertility and shorter life span* in those who are affected with schizophrenia limits the availability of potentially informative pedigrees (see Rimmer and Jacobsen 1976; Black and Fisher 1992).
6. *Characteristics of affected individuals*, such as suspiciousness and social withdrawal, may decrease the probability that they and their relatives will participate in research.
7. *The mode of inheritance is not known:* Given the lack of knowledge about how schizophrenia is transmitted (see, e.g., Risch and Baron 1984) nonparametric approaches, which do not make assumptions as to mode of inheritance, in addition to parametric linkage analyses, are being conducted routinely by many groups (Fang et al. 1995).
8. Given the probable *etiological and genetic heterogeneity* in the populations studied so far (Garver et al. 1989), very large numbers of families are necessary if there is to be any statistical power in the tests for linkage and genetic heterogeneity (Kendler and Diehl 1993).

These complications, while increasing the difficulties of the studies, are not altogether insurmountable.

Association studies. In the study of complex diseases of unknown pathophysiology, linkage studies are useful for identifying regions of interest in multiplex families. Association studies can be helpful for testing known candidate genes and for narrowing regions of interest, once they are identified. No assumptions are made about the mode of inheritance. In this strategy, association of a specific allele of a polymorphic locus is tested with the occurrence of the disease in affected individuals compared to ethnically matched controls. This strategy has a practical advantage over that of linkage because relatively scarce multiplex families are not required, and it has considerable statistical power to detect genes of "small effect" (A. Chakravarti, pers. comm.). The Transmission Disequilibrium Test (TDT) is a further refinement of the association test (Spielman et al. 1993), which tests the transmission of a particular allele from a parent to the case, using the other untransmitted allele from the parent as the "control," and this test mitigates the need for ethnically well-matched controls. Differences in the distribution of alleles between cases and controls suggest that the marker may be involved with the cause or may be in linkage disequilibrium with the cause. Several groups have reported results from such tests using polymorphisms of genes presumed to have some role in determining susceptibility to other inherited traits such as blood groups, HLA serotypes, and group-specific components (see, e.g., Fananas et al. 1990, 1992; Saha

Table 2. Association Studies Which Provide Support for Further Investigations to Determine the Possible Role These Genes May Play in Schizophrenia Susceptibility

Candidate	Covariation detected	Reference
HLA DQB1 0303		Nimgaonkar et al. (1995)
HLA DQB1 0602		
HLA class II		Zamani et al. (1994)
HLA DQB1	ethnic	Nimgaonkar et al. (1993b)
Dopamine D-2 receptor gene	age and family history	Arinami et al. (1994); Itokawa et al. (1993)[a]
Dopamine D-3 receptor gene	sex and age difference	Griffon et al. (1996); Kennedy et al. (1995)
	family history	Nimgaonkar et al. (1993a); Crocq et al. (1992)
Serontonin 2A receptor gene		Inayama et al (1996)
Neurotrophin-3 gene	sex	Dawson et al. (1995)
	severity	Hattori and Nanko (1995); Nanko et al. (1994)
Prophobilinogen deaminase gene		Sanders et al. (1992)
Tyrosine hydroxylase gene		Wei et al. (1995)

[a]Structural change in DRD2-missense nucleotide mutation from C to G.

et al. 1990). At this time, no genes have been confirmed as either contributory or noncontributory to the risk for schizophrenia. In Table 2 we summarize investigations that have provided some preliminary evidence of an association between a candidate and the susceptibility to schizophrenia (or some forms of schizophrenia). These candidates should be studied further in other populations.

Summary: Part I

How are we most likely to make progress in understanding the etiology of schizophrenia? It is by a combined epidemiological and genetic approach in which we go from one discipline to the other as each new level of information is exposed. For example, risk factors may aggregate in such a way as to suggest a subtype of schizophrenia. Then, when genes are found, their presence in families may be tested in the context of previously described subtypes, as different risk factors may be associated with different mutations. Subtypes may be further refined by additional clinical description. Finally, as a result of this alternating approach, we expect to end up with well-characterized genes whose protein products have been identified and which have been shown to be associated with predictable constellations of clinical characteristics. It should be noted that in this analysis neither epidemiology nor molecular genetics is sufficient; both are necessary and both require equal care in the collection and management of the data. For example, failure in precision in the collection of the clinical data can confound the molecular work, however carefully that work may be done, and vice versa. In the remainder of this paper we describe such an approach.

PART 2: THE JOHNS HOPKINS UNIVERSITY COLLABORATIVE SCHIZOPHRENIA STUDY

The Maryland Epidemiology Sample

The Maryland Epidemiology Sample (MES) was developed to test the hypothesis that subtypes of schizophrenia represent the outcome of different causes rather than a variable response to the same provocation. The research goal was to describe the etiologic mechanism for the different subgroups, assuming that different mechanisms would call for different treatments. A combined epidemiologic-genetic approach was developed in 1983 (see Fig. 1). Such an approach had not previously been mounted, although a similar strategy has been followed by Kendler and his colleagues studying an epidemiological sample in Ireland (Kendler et al. 1993). The need for large-scale studies was evident, given that reports of evidence of schizophrenia susceptibility were often inconsistent and seldom replicated.

By screening all psychiatric admissions to 15 facilities in the greater Baltimore area between June 1983 and April 1989, 1670 patients were chosen to participate in the studies. Patients with a broad range of psychotic diagnoses were selected for a variety of reasons, including inevitable differences in diagnostic practices across facilities and the relevance of diagnosis to genetic susceptibility. Indeed, there is evidence that susceptibility genes may be shared across different diagnostic categories (Wildenauer et al., this volume).

Detailed clinical and family history information was acquired for the patients in the MES. Prior to discharge from the hospital, an in-person diagnostic interview was administered to the patients by a clinical psychologist or social worker. The interview examined both lifetime history and current episode. Patients were also asked to allow us to (1) contact them periodically after discharge from the hospital for additional information and enrollment in further studies, (2) obtain copies of their psychiatric records, (3) talk with at least two family members to ascertain the psychiatric status of their relatives, and (4) receive copies of their birth certificate and their birth records. A concern about the accuracy of a psychiatric diagnosis based solely on information reported by the patient during the diagnostic interview led us to develop a best-estimate diagnostic procedure. This diagnosis was made by a psychiatrist who (1) reviewed all of the information reported by the patient during the diagnostic interview, (2) reviewed all of the patient's psychiatric records, and (3) spoke with the patient or an informant 6 months after the patient's discharge from the hospital to obtain course and outcome information. The diagnosis was made according to DSM-III criteria.

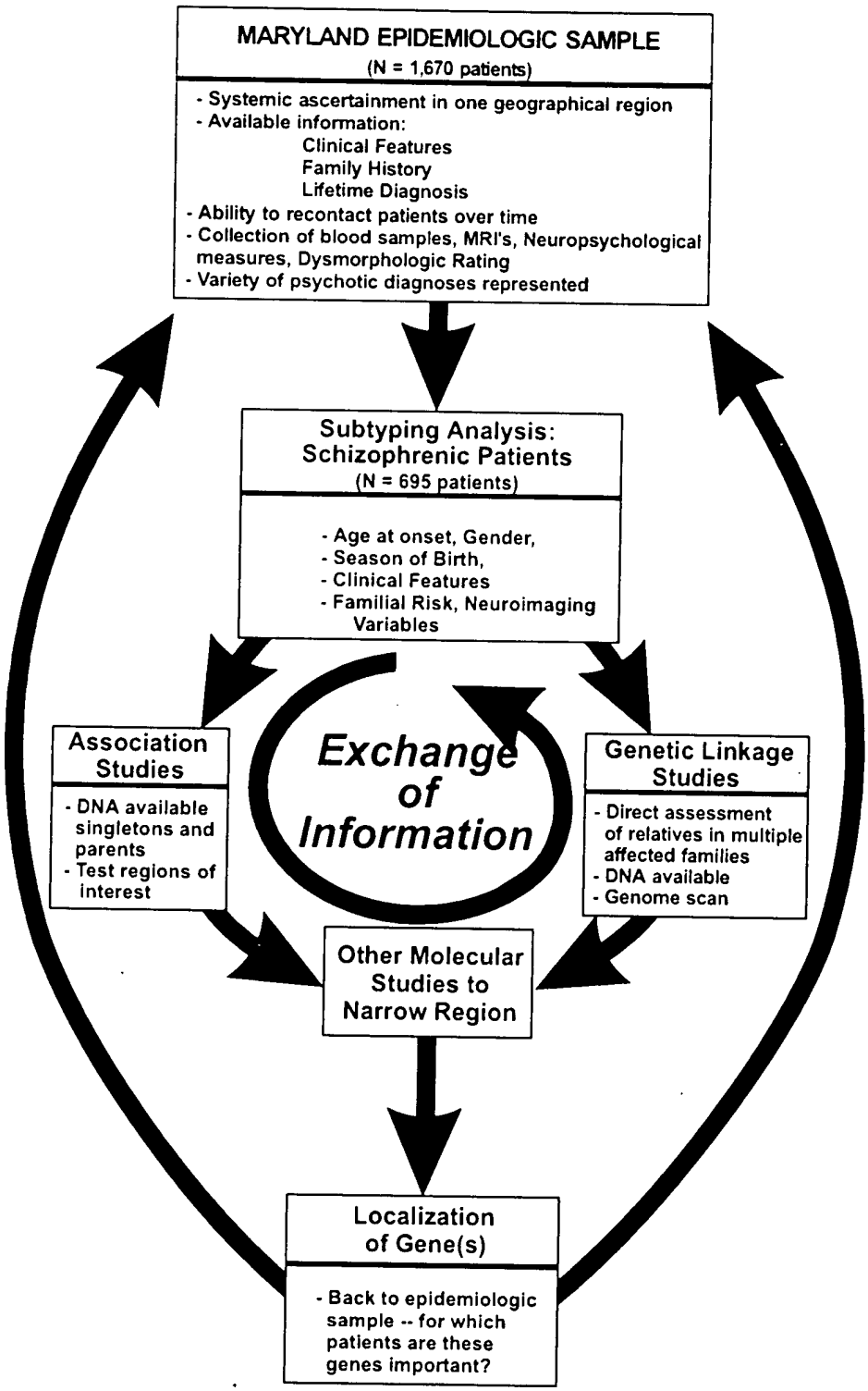


Figure 1. Schematic of epidemiologic/genetic approach.

For consenting patients in the MES (77% of the patients), a family history was obtained from two informants judged by the patient to be knowledgeable about his or her family. During a telephone interview, the informants were systematically questioned about the medical and psychiatric histories of the patient's first- and second-degree relatives. We used modified Family History Research Diagnostic Criteria to classify

the relatives (Endicott et al. 1975). Since the family history method has been shown to be less accurate than directly assessing individuals for the identification of schizophrenia (Andreasen et al. 1977), we considered these data as preliminary to the collection of the data necessary to classify the relatives of probands for molecular studies. However, the family history data were used to determine whether patient characteristics delineate subtypes of schizophrenia that may differ in their etiologies. The results of these subtyping analyses are presented below.

Results of Initial Subtyping Analyses

The initial subtyping factors of interest were gender, age at onset, season of birth, and family history of schizophrenia. Utilizing family history information about the patients in the MES with a best-estimate diagnosis of schizophrenia, we have demonstrated the following: (1) first-degree relatives of female schizophrenics have a greater risk of developing schizophrenia than do first-degree relatives of male schizophrenics (Wolyniec et al. 1992); (2) first-degree relatives of schizophrenics with an early onset of illness (less than age 17) are at greater risk of developing schizophrenia than are first-degree relatives of schizophrenics with a later onset of illness (Pulver et al. 1990; Boyle et al. 1991); (3) first-degree relatives of schizophrenics born during the winter–spring months (specifically, February through May) are at higher risk of developing schizophrenia than are relatives of schizophrenics born during the other months of the year (Pulver et al. 1992b); (4) mothers of winter–spring-born schizophrenics have a greater spacing between their live births than do mothers of winter–spring-born controls (Pulver et al. 1992a); (5) the first-degree relatives of schizophrenic patients who also have a history of manic symptoms are at an increased risk for developing schizophrenia (Liang and Pulver 1996) than are the relatives of schizophrenic patients who do not have manic symptoms, and (6) the first-degree relatives of schizophrenic patients who also have a history of obsessive and compulsive symptoms are at a greater risk for depression and suicide than are the relatives of schizophrenics who do not have obsessive and compulsive symptoms (Samuels et al., in prep.).

On the basis of these results, we suggest that some individuals at an increased risk for schizophrenia due to genetic background may also be more vulnerable to some seasonally varying factor which further increases their risk (e.g., viral exposure in utero, pregnancy and birth complications). The greater spacing between the births of the mothers of winter–spring-born schizophrenic patients suggests that there might be other complicated pregnancies resulting in a nonviable outcome. It is important to determine whether the clinical features of schizophrenia which are associated with familial risk (i.e., age at onset, manic symptoms, obsessive-compulsive symptoms) characterize families, or some subset of families, or sibships. If this is the case, it will provide additional support for the idea that schizophrenia is etiologically heterogeneous, and these particular covariates may be helpful in discriminating between subgroups of patients in molecular studies.

Genetic Linkage Study

In 1989, a second phase of the work began. Here, the goal was to ascertain a set of families in which more than one individual was affected with schizophrenia (multiplex families) to be used to conduct a genome-wide search for genes and DNA markers associated with the risk for schizophrenia. The first source of such families was the MES. The families of schizophrenic patients who, on the basis of family history information, had at least one other living first- or second-degree relative with a diagnosis of either schizophrenia, schizoaffective disorder, or unspecified functional psychosis were considered eligible for this study. Less than one third of the schizophrenic patients in the MES had one or more affected relatives (for a description of the occurrence of schizophrenia in the relatives, see Pulver and Bale 1989). If direct assessment of the second affected family member confirmed a diagnosis of schizophrenia or schizoaffective disorder, complete ascertainment of the family was pursued. Complete ascertainment included direct psychiatric examinations and blood drawings for all of the proband's first- and second-degree relatives, as well as all of the first-degree relatives of other affected individuals in the family. Spouses of affected individuals were also included. Because of the need for greater statistical power in linkage studies, additional multiplex schizophrenic families were also recruited through physician referrals. The nonsystematically ascertained families were more likely to have at least three affected individuals.

Design features of the linkage study. It is our belief that quality control measures must be built into all aspects of data collection and analysis in a study of this kind. Some of the key design features of the linkage study include:

1. The majority of the families in the set used for the genome-wide scan were ascertained through a systematically selected group of patients (the MES). The set of families can be characterized by ascertainment class—either systematic or opportunistic.
2. The ascertainment strategy was specified in advance of data collection. We attempted to acquire information about all first- and second-degree relatives of affected individuals in each family. This information was gathered through any or all of the following: reports from family informants, review of psychiatric treatment records, direct psychiatric ex-

aminations, and reports from treating therapists and psychiatrists.

3. Rules for extending the pedigrees were specified; pedigrees were extended through affected individuals.
4. Direct psychiatric examinations of relatives were made by clinical psychologists or psychiatrists using standard psychiatric assessment instruments. The psychiatric assessments were designed to ascertain symptoms relevant to major mental disorders, as well as severe personality disorders. Information reported by the subject was supplemented by interviewing an informant identified by the subject. Examinations were audiotaped with the permission of the subject for quality control purposes.
5. To increase the accuracy of the assignment of affected status, final diagnoses were formulated through a consensus procedure. For each subject, two psychiatrists or psychologists independently reviewed all available information including the diagnostic interview, information reported by informants, psychiatric treatment records, and information reported by the subject's treating physician, and, after review, completed a DSM-III-R diagnostic assignment. If the two assignments agreed, the diagnosis was considered final. If there was disagreement, the case was discussed by the two raters until a consensus was reached. If a consensus could not be reached, additional information was sought and/or a third diagnostician was asked to review the case and to try to resolve the discrepancies. In some instances, this was not possible and the case was then classified as affected status unknown. Consensus was necessary not only for the diagnoses assigned, but also for age-at-onset ratings for each diagnosis.
6. The definition of affected and unknown phenotypes was specified prior to the analysis. This was done so that specific hypotheses could be tested rather than fishing for a positive result. For the analyses in the genome-wide search, individuals with a consensus diagnosis of either schizophrenia or schizoaffective disorder were considered affected; individuals with other psychotic disorders or schizotypal, schizoid, or paranoid personality disorder were classified as unknown.
7. To control for possible bias in coding genotyping data, molecular genetics laboratory personnel were blind to the affected status of all DNA samples in the laboratory, and all genotyping data were coded after consensus of two readers. Similarly, the diagnosticians were blind to the genotyping data. In addition, the genetic analysts were not involved in clinical or molecular assessments.
8. Given the variable age at onset of many psychiatric diseases, and the fact that many relatives have not lived through all the known risk periods for disease, a longitudinal study design was implemented to allow us to maintain current affected status of all individuals in the pedigree through regularly updated interviews.

Table 3. Description of 54 Pedigrees

Total number of individuals	456	
individuals genotyped	363	(80%)
Affected individuals	147	
age at onset		
<35	125	(85%)
≥35	17	(15%)
average	24.6	
sex		
male	90	(61%)
female	57	(39%)
DSM-III-R diagnosis		
schizophrenia	143	(97%)
schizoaffective	4	(3%)
average number of affecteds in each family	2.7	
Unaffected individuals	246	
age at assessment		
<35	30	(12%)
≥35	226	(88%)
Individuals with diagnosis "unknown"	63	
Linearity in pedigrees (re: proband)		
maternal	19	(35%)
paternal	8	(15%)
sibs only	27	(50%)

Description of the families being genotyped for the genome-wide scan. At this time, we have completed the diagnostic evaluations of more than 80 families. For the past 18 months, we have been genotyping 363 individuals in 54 pedigrees. Additional pedigrees are being added as we continue the research. However, at this time, all analyses are based on 54 families. Characteristics of the families are summarized in Table 3.

Strategy for the genome-wide search. Given the likelihood that no one gene will have a sufficient effect in all of the families, simulation studies revealed that the number of families we had available was insufficient to detect linkage for a gene contributing to liability in only a small proportion of families (under heterogeneity). We embarked on our studies knowing that collaborative efforts with other research groups were going to be necessary to detect genes that may be important to subsets of the schizophrenic families. Therefore, a three-stage strategy was developed; the first two stages involved typing markers in the families we ascertained, the third stage was to help facilitate the study of the regions that met criteria in other family samples (by other investigators) (see Table 4).

In analyses of our 54 families, low positive LOD scores were used as thresholds to detect regions of interest from a sparse linkage map. Additional markers were typed to create a dense map only at promising genomic regions. The staged approach was used to minimize the number of markers required for typing. An initial critical LOD score for stage I was selected that was highly likely to include any true linkage (LOD≥ 0.30). The strategy was to space markers every

Table 4. Three-stage Searching Strategy

Stage 1.	2-Point linkage analyses 54 JHU families Find markers where LRH ≥ 2.0 (LOD ≥ 0.30)
Stage 2.	2-Point linkage analyses 54 JHU families Find markers where LRH ≥ 10.0 (LOD ≥ 2.0)
Stage 3.	"Hot spot" Facilitate follow-up of region by other investigators in independent families

20 cM, adding additional markers (every 3 cM) to any selected regions of 40 cM. The critical LOD score for stage II was selected to be greater than or equal to 2.0. Any regions meeting the stage II criterion were then released to other investigators for independent replication (stage III).

Statistical analysis. Affected subjects for all statistical analyses were restricted to those satisfying DSM-III-R (American Psychiatric Association 1987) criteria for schizophrenia or schizoaffective disorder. Subjects with other psychotic disorders or "schizophrenia spectrum personality disorders" were classified as "unknown" phenotype with unknown liability. Both autosomal dominant and recessive models were used. Given that the concordance rate for schizophrenia among monozygotic twins has been reported to be between 50% and 65%, the classification of those who are unaffected is problematic. Therefore, the two parametric models are defined with "affecteds only" liability classes, where unaffected subjects were treated the same as unknown subjects; their genotypings are available, but their phenotype has no effect on the LOD score. In addition to linkage analyses using two models, nonparametric sib-pair analyses to test for sharing of alleles identical-by-descent were conducted. However, the results from linkage analyses using the autosomal dominant and recessive models were used to guide the laboratory through the two-stage strategy.

The programs MLINK, LODSCORE, ILINK (Lathrop and Lalouel 1984; Lathrop et al. 1984, 1986), and HOMOG (Ott 1991) were used, via the FASTLINK versions (Cottingham et al. 1993; Schaffer et al. 1994) as implemented in the ANALYZE software (Terwilliger 1994) to compute 2-point maximum LOD scores under heterogeneity, for each locus tested. The ANALYZE software also estimates the proportion of alleles shared identical-by-descent in all available sib pairs (Kuokkanen et al. 1996; Satsangi et al. 1996). In addition, within the past several months, a novel and efficient way to run nonparametric multipoint analyses for many markers in a dense map has become available. Some of the data have been analyzed with these programs, known as MAPMAKER/SIBS (Kruglyak and Lander 1995) and GENEHUNTER (Kruglyak et al. 1996).

Results from the Genome Scan

We have typed and analyzed more than 650 markers, on all 22 autosomes, in the genome scan. We approximate that these markers cover 70% of the genome. Using the staged strategy, we have identified four regions which we felt were worthy of follow-up. These regions are 22q11-q13; 8p22- p21; 3p26-p24; and 13q14-q32. The results of analyses for each of these regions, as well as the data reported from other groups for these regions, are summarized below. In addition, we report the results of analyses of data generated to follow up a report from Straub et al. (1995), suggesting linkage on chromosome 6p24-p22.

Chromosome 22q. The region of interest on chromosome 22 is shown in Figure 2 (Pulver et al. 1994b). This is the first region identified in our genome scan, satisfying criteria for stage II. The maximum LOD score was obtained for the dominant model (LOD = 2.05) for the marker *D22S445* in 22q13. The highest LOD score obtained using the recessive model was 1.39 for *D22S685* mapped to 22q12. The most significant sib-pair sharing was found for the markers *D22S1268, D22S686* (p=.007) in 22q12-q13. The region

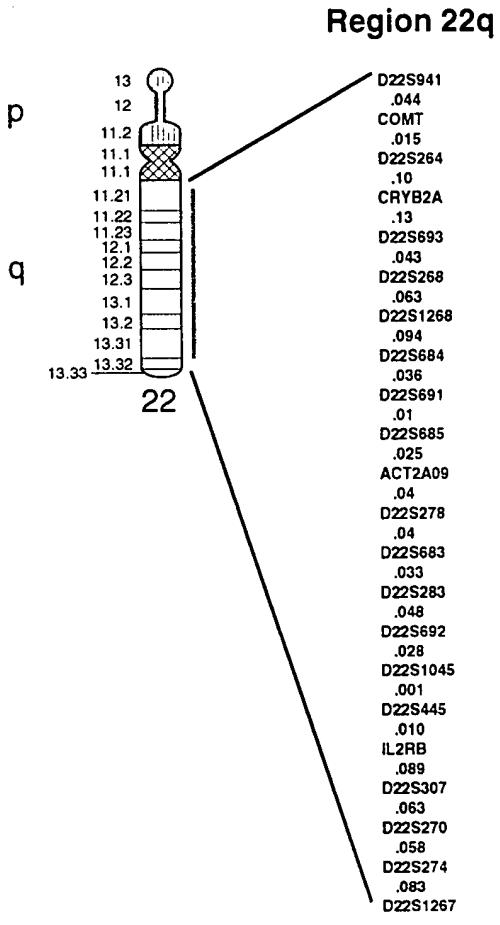


Figure 2. Ideogram of chromosome 22.

of allele sharing is relatively large. These data alone provide only suggestive evidence for a susceptibility gene somewhere in this region on chromosome 22. Recently completed multipoint analyses confirm that the region of interest is large and may, in fact, contain two separate regions of interest.

Prior to publishing the initial suggestive findings for this region, we organized a multicenter collaborative effort to investigate the region further. This was the first such effort to be taken in this field. Previous experience with a report "Localization of a Susceptibility Locus for Schizophrenia on Chromosome 5" (Sherrington et al. 1988), followed by a series of papers reporting no evidence for linkage (Kennedy et al. 1988; St. Clair et al. 1989; Aschauer et al. 1990; McGuffin et al. 1990; Crowe et al. 1991; Campion et al. 1992; Macciardi et al. 1992), was problematic for the field of psychiatric genetics (see, e.g., Owen 1992). The early pronouncement of a finding of a schizophrenia susceptibility gene followed by many failures to replicate caused disillusionment with the scientific goal and adversely affected the reputation of psychiatric genetics among the general scientific community. Prior to publication of our results on 22q (Pulver et al. 1994b), three research groups known to be genotyping families with schizophrenia were contacted in 1992 and advised of the findings; i.e., Kenneth Kendler and his colleagues, Michael Owen and Michael Gill and their colleagues, and Jacque Mallet and his colleagues. A collaboration was begun; interlaboratory reliability exercises were conducted for both the genotyping data and the diagnostic assignments.

As a result of the collaborative study, 217 additional families were genotyped for three markers on 22q (*D22S268*, *IL2RB*, and *D22S307*). The LOD score analyses for this independent sample of families did not support linkage (Pulver et al. 1994c). However, since power calculations showed that there was low power to detect linkage if 25% or less of the families were linked to markers in this region, we concluded that additional efforts needed to be mounted to test the hypothesis that there was a schizophrenia susceptibility gene in this region in some families. After we made our findings public, Coon and colleagues notified us that as part of their genome search using 9 families, they detected a LOD score of 2.09 for a recessive model toward the distal end of the chromosome at *D22S276*. Subsequently, their data were published (Coon et al. 1994a). Since then, several additional groups genotyped markers in the region. Although all of the reports have not supported linkage, none of them has been able to rule out the possibility of linkage for markers in this region for some subset of the schizophrenic families. Additional support has accrued: (1) Gill et al. (1996) reported their best evidence at *D22S278* in 22q12 ($p<.01$); (2) Schwab et al. (1995b) reported a mildly positive LOD (0.61) for *D22S304* which is approximately 3.5 cm centromeric to *D22S278*; (3) Polymeropoulos et al. (1994) reported greater than expected allele sharing for several markers in the region around *D22S278* ($p<.02$); (4) Moises et al. (1995a) reported some evidence for linkage disequilibrium between schizophrenia and allele #243 of *D22S278* ($p=0.02$) in 113 unrelated schizophrenic subjects and their 226 parents; and finally (5) a collaborative effort organized by Michael Gill and Homero Vallada et al. (Gill et al. 1996) employing an affected sib-pair analysis analyzed genotypings of one marker (*D22S278* in 22q12) typed by 11 research groups. From a pooled sample of more than 600 families, the analysis of the sub sample of 296 affected sib-pairs for whom parental data were available revealed a significant excess of alleles shared by affected individuals for the marker *D22S278* ($p=.001$).

In addition to the evidence for the region around *D22S278*, there is some evidence for a schizophrenia-related susceptibility locus centromeric to this region. The evidence is as follows: (1) Although for some time it had been observed that children with velo-cardio facial syndrome (VCFS), a congenital disorder associated with an interstitial deletion of 22q11, have a range of psychiatric and learning problems, a study using appropriate research methods found that by examining the psychiatric status of VCFS subjects who were between the ages of 16 and 25, a higher than expected rate of schizophrenia was found (Pulver et al. 1994a). (2) On the basis of the evaluation of 100 schizophrenic patients ascertained from the MES, we have reported a higher rate of the 22q11 deletion among the schizophrenics than would be expected based on estimated population prevalence (Karayiorgou et al. 1995; Lindsay et al. 1995). Details of this work as well as the results of additional work in this region are reported by Karayiorgou et al. (this volume).

Chromosome 8p22-p12. The region of interest on chromosome 8 (Pulver et al. 1995) is shown in Figure 3. The dominant model provided a maximum LOD score in this region for the marker *SFTP2* (2.40) in 8p21. The maximum LOD score for the recessive model was 2.10 for the marker *D8S136*, tightly linked to *SFTP2*. The most significant evidence for a susceptibility locus in this region from the analyses of allele sharing among sib pairs is for the marker *D8S258* ($p<.001$). Results of multipoint sib-pair analyses recently completed provide additional support for this region. A report of these analyses is currently being prepared for publication.

Given the strength and the consistency of these data, we requested that colleagues again join in a collaborative effort to study this region (stage III of our strategy). Douglas Levinson agreed to coordinate a study of 8p (and two other regions) using family data from 14 groups with a total of 567 informative pedigrees. The results of the analyses of five markers on 8p (*D8S261, D8S258, D8133, D8136,* and *D8S283)* have been completed and provide additional support for the presence of a schizophrenia susceptibility gene

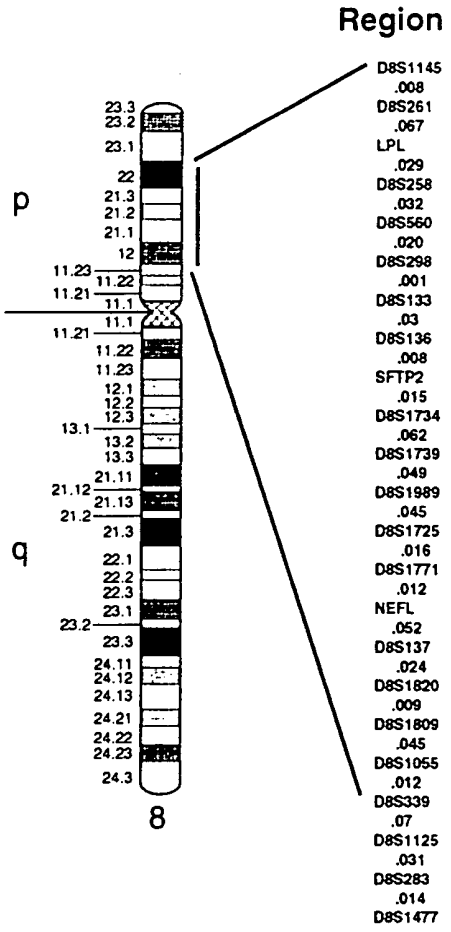


Figure 3. Ideogram of chromosome 8.

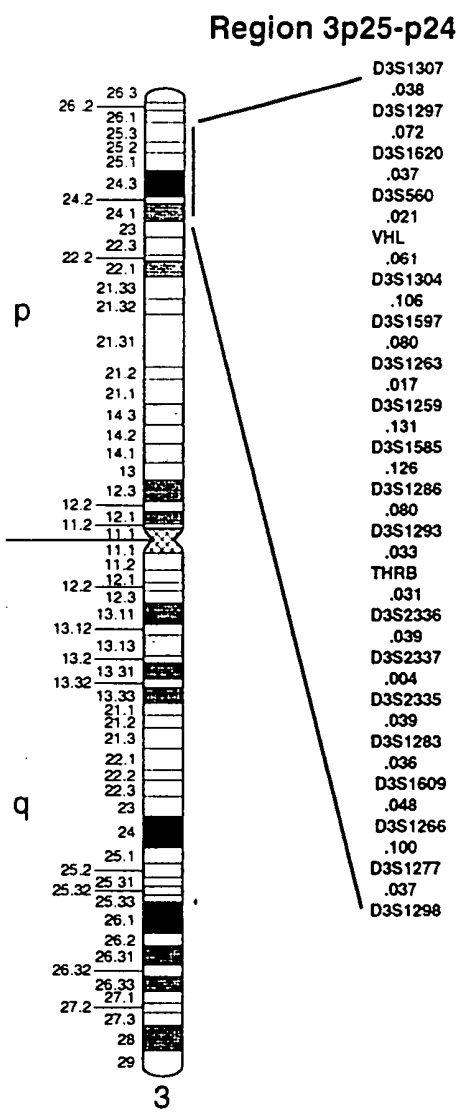


Figure 4. Ideogram of chromosome 3.

in this region (D. Levinson et al.; Schizophrenia Collaborative Linkage Group; both in prep.). The maximum LOD score 2.22 (p=.0014) for 13 groups (without our data) was obtained with a recessive model for locus *D8S261*, the most telomeric of the five markers tested. The nonparametric analyses report p<.005 for *D8S133* for allele sharing in the independent replication sample. Slightly positive LODs (>0.3) were observed in 10 of 13 groups.

Chromosome 3p26-p24. Figure 4 shows the region 3p26-p24, a region of approximately 60 cM mapped by 22-dinucleotide-repeat polymorphisms. The maximum LOD score in this region was obtained with an "affecteds only" dominant model for the marker *D3S1283* (LOD score 1.56). The maximum LOD score in this region using the affecteds-only recessive model was also at *D3S1283* (LOD score 1.75). Reanalyses with our current sample, and additional markers, now show that only one marker, *D3S1283*, shows evidence (p=.02) of excess sib-pair sharing. Markers in this region of chromosome 3 were also genotyped in the replication effort coordinated by Levinson. All heterogeneity LODs for four markers used in the follow-up effort *(D3S1293, D3S1283, D3S1266, D3S1298)* were less than 0.08 and the region was therefore considered negative. However, these data do not provide evidence to exclude the possibility of a schizophrenia susceptibility gene in the region in some individuals.

Chromosome 13q14-q32. The region of interest on the chromosome is 13q14-q32 and is shown in Figure 5. The greatest evidence for this region comes from the sib-pair analyses for the marker *D13S770* (p=.0002). The maximum LOD score in this region was obtained with a dominant model for the marker *D13S128* (LOD = 3.24). The maximum LOD score for the recessive model was 2.53 for the marker *D13S779*. Recently completed multipoint sib-pair sharing analyses for this region add support for a schizophrenia susceptibility locus in this region. A more dense map of this region is currently being genotyped.

Other evidence to support the existence of a susceptibility gene for schizophrenia in region 13q had pre-

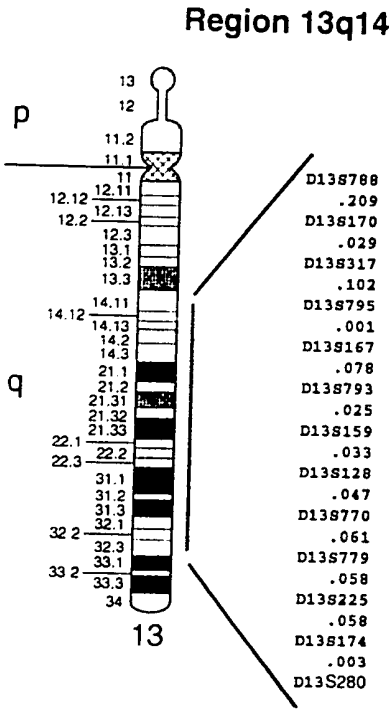


Figure 5. Ideogram of chromosome 13.

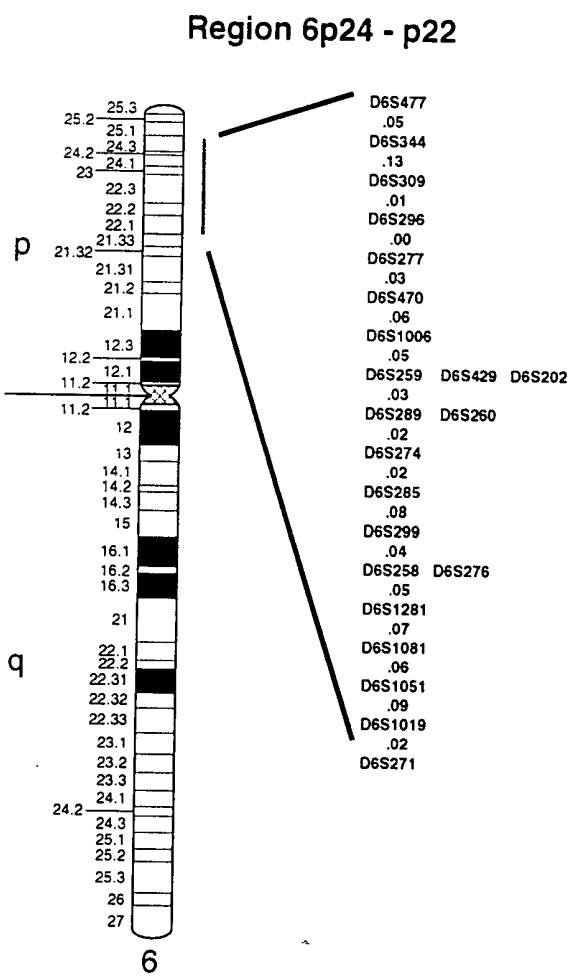


Figure 6. Ideogram of chromosome 6.

viously been reported by Lin et al. (1995) in their study of 13 schizophrenia pedigrees in the region 13q14.1-q32. LOD score analysis of markers in this region was moderately positive (*D13S144;* LOD=1.61). Multipoint results from a dominant model allowing for heterogeneity using a narrow definition of affected gave a maximum multipoint LOD score of 2.00 for *D13S128* in 13q32.

Chromosome 6p. As part of the 3-stage strategy, we also follow up reports of regions of interest from other groups. Kendler et al. informed us of their finding of a maximum LOD score of 3.51 for *D6S296* in their large Irish sample of nearly 300 pedigrees (Straub et al. 1995). There was also suggestive but weaker evidence for sib-pair sharing of alleles at *D6S296*. We have followed up this suggestion of linkage by typing markers in both 6p24-22 and in the more centromeric HLA region (6p21) (Antonarakis et al. 1995) (see Fig. 6). The strongest sib-pair sharing in the sample was seen for marker *D6S296* (p=.0003). The marker *D6S1006* gave the highest LOD for the dominant model (0.96) and also gave the highest recessive LOD (1.32).

Additional evidence supporting the existence of a schizophrenia susceptibility has also been obtained in several samples (Gurling et al. 1995; Schwab et al. 1995a; Wang et al. 1995) and the large collaborative effort organized by Levinson et al. (1996). These results of the analysis of the collaborative effort as well as additional work in this region by the Kendler group are reported by Straub et al. (this volume).

Summary: Part II

The results of our epidemiologic studies support the hypothesis that schizophrenia is etiologically heterogeneous. The true test of heterogeneity will only be possible when the specific etiologic mechanism(s) is delineated. The identification of schizophrenia susceptibility genes is an important step toward attaining this goal.

Over the last several years, significant progress has been made toward elucidating the genes associated with schizophrenia susceptibility. Our group has now reported four regions of interest from the genome scan (i.e., chromosomes 22q 11-13, 8p22-p12, 3p26-24, and 13q14-q32). Multicenter follow-up studies have now been completed for three of the regions (22q, 8p, and 3p), and evidence to support linkage has been obtained for two of the regions (22q and 8p). The region on 3p, although studied in a large collaborative follow-up study, did not provide evidence supporting linkage. This paper represents our first report of the region on chromosome 13q. We are hopeful that a collaborative follow-up will be organized to test this region. In addition to these regions, the cumulative evidence now

available for the region of interest identified by Kendler and colleagues (Straub et al. 1995) on chromosome 6p24-p22 provides support for the existence of a susceptibility gene in this region as well.

In all likelihood, over the next several years we will learn of the existence of other regions of interest and other potential candidate genes. However, the identification of regions of interest is an initial step toward the next goal: the identification of the gene(s). The problem that we and others are now struggling with is how to move from a region of interest or a candidate region to the identification of a gene. Our group has articulated a plan, described below, to narrow the regions of interest. This plan requires an additional huge effort from scientists representing all of the disciplines in our multidisciplinary team.

Association/linkage disequilibrium tests. As a first step to follow up linkage results in family samples, association studies are conducted using a new sample of schizophrenic patients and their parents from simplex families in the epidemiologic sample; these families have not previously been analyzed in linkage studies. We are also planning a new data collection effort; i.e., to ascertain a sample of Ashkenazi Jewish schizophrenic patients and their families for both sib-pair studies and association studies. This new sample will be tested to determine if the previously identified regions of interest are important in this more homogeneous group, as well as to identify new regions of interest. In regions where the TDT and/or association studies confirm disequilibrium, family material can be used for molecular cloning strategies.

Refinement of disease definition/heterogeneity analyses. All of the analyses that we have reported thus far have been based on a strict definition of affected; i.e., a consensus diagnosis of schizophrenia or schizoaffective disorder. Additional analyses are planned that will expand the definition of affected to include what has previously been described as schizophrenia spectrum disease (schizotypal, schizoid, and paranoid personality disorders and unspecified psychotic disorders).

It has long been recognized that there is considerable clinical heterogeneity in schizophrenia. Recently, several independent studies employing multivariate techniques such as factor analysis and latent class analysis have consistently suggested that the symptoms of schizophrenia can be grouped into three dimensions: negative symptoms (affective flattening and negative thought disorder), disorganization (inappropriate affect and positive thought disorder), and positive psychotic symptoms (e.g., hallucinations) (for review, see Andreasen et al. 1995). The consistency of the findings across studies, and the presumed etiologic heterogeneity in schizophrenia, suggest that these dimensions may be good candidates to be used in the definition of affected in linkage studies. It may be that genetic heterogeneity is tied to the clinical heterogeneity that has been observed. We are currently in the process of (1) confirming that these clinical dimensions identified by others can be observed in the families and (2) determining how classification of affected individuals according to these dimensions influences linkage results. Multipoint analyses in the regions of interest will be conducted treating these clinical dimensions, as well as other variables such as season of birth and age at onset, as covariates.

Molecular analyses. In addition to what is described above, the molecular biologists have the following work ongoing or planned: (1) identify any potential candidate gene in narrowed regions and screen for mutations, (2) develop a yeast artificial chromosome (YAC) contig across the narrowed regions, (3) subclone YACs into cosmids, (4) identify genes through either exon trapping or cDNA screening, (5) screen genes for mutations, and (6) sequence the variations.

The goals which follow gene identification are to understand the gene function, to describe the mechanism responsible for the disease, to test the heterogeneity hypothesis, and to describe more etiologically homogeneous groups. Of course, it is our hope that success in this work will lead to the development of a new understanding of schizophrenia which will allow the development of new, more effective medications.

ACKNOWLEDGMENTS

The authors acknowledge the contribution of the patients and their families, Dr. William T. Carpenter, Jr. and the Maryland Psychiatric Research Center, Drs. Jurg Ott, Aravinda Chakravarti, and Swapan Nath. In addition, we are appreciative of the financial support obtained from Sandoz Pharmaceuticals and National Institute of Mental Health grants 1 R01 MH-45588 and 1 R01 35712 and National Institutes of Health grant DRR-OPD-GCRC #RR00722.

REFERENCES

American Psychiatric Association. 1987. *Diagnostic and statistical manual of mental disorders,* 3rd edition, revised. American Psychiatric Association, Washington, D.C.

———. 1994. *Diagnostic and statistical manual of mental disorders,* 4th edition. American Psychiatric Association, Washington, D.C.

Andreasen, N.C., J. Endicott, R.L. Spitzer, and G. Winokur. 1977. The family history method using diagnostic criteria. Reliability and validity. *Arch. Gen. Psychiatry* **34:** 1229.

Andreasen, N.C., S. Arndt, R. Alliger, D. Miller, and M. Flaum. 1995. Symptoms of schizophrenia. *Arch. Gen. Psychiatry* **52:** 341.

Antonarakis, S.E., J.L. Blouin, A.E. Pulver, P. Wolyniec, V.K. Lasseter, G. Nestadt, L. Kasch, R. Babb, H.H. Kazazian, B. Dombroski, et al. 1995. Schizophrenia susceptibility and chromosme 6p24-22. *Nat. Genet.* **11:** 235.

Arinami, T., M. Itokawa, H. Enguchi, H. Tagaya, S. Yano, H. Shimizu, H. Hamaguchi, and M. Toru. 1994. Association of dopamine D2 receptor molecular variant with

schizophrenia (comments). *Lancet* **343:** 703.

Aschauer, H.N., G. Aschauer-Treiber, K.E. Isenberg, R.D. Todd, M.A. Knesevich, D.L. Garver, T. Reich, and C.R. Cloninger. 1990. No evidence for linkage between chromosome 5 markers and schizophrenia. *Hum. Hered.* **40:** 109.

Asherson, P., E. Parfitt, M. Sargeant, S. Tidmarsh, P. Buckland, C. Taylor, A. Clements, M. Gill, P. McGuffin, and M. Owen. 1992. No evidence for a pseudoautosomal locus in schizophrenia linkage analysis of mulitply affected families. *Br. J. Psychiatry* **161:** 63.

Barr, C.L., J.L. Kennedy, A.J. Pakstis, C.M. Castiglione, J.R. Kidd, L. Wetterberg, and K.K. Kidd. 1994. Linkage study of a susceptibility locus for schizophrenia in the pseudoautosomal region. *Schizophr. Bull.* **20:** 277.

Barr, C.L., J.L. Kennedy, A.J. Pakstis, L. Wetterburg, S.L. Beirut, C. Wadelius, J. Wahlstrom, L. Giuffra, L. Martinsson, J. Gelernter, J. Hallmayer, H.W. Moises, J. Kurth, L.L. Cavalli-Sforza, and K.K. Kidd. 1994. Progress in a genome scan for linkage in schizophrenia in a large Swedish kindred. *Am. J. Med. Genet.* **54:** 51.

Bassett, A.S. 1992. Chromosomal aberrations and schizophrenia autosomes. *Br. J. Psychiatry* **161:** 323.

Benes, F.M. 1993. Neurobiological investigations in cingulate cortex of schizophrenic brain. *Schizophr. Bull.* **19:** 537.

Benes, F.M., J. Davidson, and E.D. Bird. 1986. Quantitative cytoarchitectural studies of the cerebral cortex of schizophrenics. *Arch. Gen. Psychiatry* **43:** 31.

Black, D.W. and R. Fisher. 1992. Mortality in DSM-IIIR schizophrenia. *Schizophr. Res.* **7:** 109.

Bland, R.C. 1982. Predicting the outcome in schizophrenia. *Can. J. Psychiatry* **27:** 52.

Boyd, J.H., A. E. Pulver, and W. Stewart. 1986. Season of birth: Schizophrenia and bipolar disorder. *Schizophr. Bull.* **12:** 173.

Boyle, M.H., D.R. Offord, Y.A. Racine, and G. Catlin. 1991. Ontario child health study follow-up: Evaluation of sample loss. *J. Am. Acad. Child Adolesc. Psychiatry* **30:** 449.

Campion, D., T. D'Amato, H. Laklou, O. Sabate, M. Jay, M. Leboyer, A. Malafosse, P. Gorwood, M.C. Babron, D. Hillaire, F. Clerget-Darpoux, G. Waksman, and J. Mallet. 1992. Failure to replicate linkage between chromosome 5q11-q13 markers and schizophrenia in 28 families. *Psychiatry Res.* **4:** 171.

Cannon, T.D. and E. Marco. 1994. Structural brain abnormalities as indicators of vulnerability to schizophrenia. *Schizophr. Bull.* **20:** 89.

Cannon-Spoor, H.E., S.G. Potkin, and R.J. Wyatt. 1982. Measurement of premorbid adjustment in chronic schizophrenia. *Schizophr. Bull.* **8:** 470.

Collinge, J., L.E. Delisi, A. Boccio, E.C. Johnstone, A. Lane, C. Larkin, M. Leach, R. Lofthouse, F. Owen, M. Poulter, T. Shah, C. Walsh, and T.J. Crow. 1991. Evidence for a pseudo-autosomal locus for schizophrenia using the method of affected sibling pairs. *Br. J. Psychiatry.* **158:** 624.

Coon, H., J. Holik, M. Hoff, F. Reimherr, P. Wender, M. Myles-Worsley, M. Waldo, R. Freedman, and W. Byerley. 1994a. Analysis of chromosome 22 markers in nine schizophrenia pedigrees. *Am. J. Med. Genet.* **54:** 72.

Coon, H., S. Jensen, J. Holik, M. Hoff, M. Myles-Worsley, F. Reimherr, P. Wender, M. Waldo, R. Freedman, M. Leppert, and W. Byerley. 1994b. Genomic scan for genes predisposing to schizophrenia. *Am. J. Med. Genet.* **54:** 59.

Cottingham, R.W., Jr., R.M. Idury, and A. A. Schaffer. 1993. Faster sequential genetic linkage computations. *Am. J. Hum. Genet.* **53:** 252.

Crocq, M.A., R. Mant, P. Asherson, J. Williams, Y. Hode, A. Mayerova, D. Collier, L. Lannfelt, P. Sokoloff, J.C. Schwartz, M. Gill, J.P. Macher, P. McGuffin, and M.J. Owen. 1992. Association between schizophrenia and homozygosity at the dopamine D3 receptor gene (comments). *J. Med. Genet.* **29:** 858.

Crow, T.J., L.E. Delisi, R. Lofthouse, M. Poulter, T. Lehner, N. Bass, T. Shah, C. Walsh, A. Boccio-Smith, G. Shields, and J. Ott. 1994. An examination of linkage of schizophrenia and schizoaffective disorder to the pseudoautosomal region (Xp22.3) (comments). *Br. J. Psychiatry* **164:** 159.

Crowe, R.R., D.W. Black, R. Wesner, N.C. Andreasen, A. Cookman, and J. Roby. 1991. Lack of linkage to chromosome 5q11-q13 markers in six schizophrenia pedigrees. *Arch. Gen. Psychiatry* **48:** 357.

D'Amato, T., G. Waksman, M. Martinez, C. Laurent, P. Gorwood, D. Campion, M. Jay, C. Petit, C. Savoye, C. Bastard, M.C. Babron, F. Clerget-Darpoux, and J. Mallet. 1994. Pseudoautosomal region in schizophrenia: Linkage analysis of seven loci by sib-pair and lod-score methods. *Psychiatry Res.* **52:** 135.

D'Amato, T., D. Campion, P. Gorwood, M. Jay, O. Sabate, C. Petit, M. Abbar, A. Malafosse, M. Leboyer, D. Hillaire, F. Clerget-Darpoux, J. Feingold, G. Waksman, and J. Mallet. 1992. Evidence for a pseudoautosomal locus for schizophrenia. II. Replication of a non-random segregation of alleles at the DXYS14 locus. *Br. J. Psychiatry.* **161:** 59.

Dawson, E., J.F. Powell, P.C. Sham, M. Nothen, M.A. Crocq, P. Propping, J. Korner, M. Rietschel, J. Vanos, P. Wright, R.M. Murray, and M. Gill. 1995. An association study of a neurotrophin-3 (nt-3) gene polymorphism with schizophrenia. *Acta Psychiatr. Scand.* **92:** 425.

Delisi, L.E., U. Friedrich, J. Wahlstrom, A. Boccio-Smith, A. Forsman, K. Eklund, and T.J. Crow. 1994a. Schizophrenia and sex chromosome anomalies. *Schizophr. Bull.* **20:** 495.

Delisi, L.E., M. Devoto, R. Lofthouse, M. Poulter, A. Smith, G. Shields, N. Bass, G. Chen, A. Vita, C. Morganti, J. Ott, and T.J. Crow. 1994b. Search for linkage to schizophrenia on the X and Y chromosomes. *Am. J. Med. Genet.* **54:** 113.

Devlin, B. and N. Risch. 1995. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* **29:** 311.

Done, D.J., T.J. Crow, E.C. Johnstone, and A. Sacker. 1994. Childhood antecedents of schizophrenia and effective illness: Social adjustment at ages 7 and 11. *Br. Med. J.* **309:** 699.

Eaton, W.W. 1985. Epidemiology of schizophrenia. *Epidemiol. Rev.* **7:** 105.

———. 1991. Update on the epidemiology of schizophrenia. *Epidemiol. Rev.* **13:** 320.

Eaton, W.W., C. Hayward, and R. Ram. 1992. Schizophrenia and rheumatoid arthritis: A review. *Schizophr. Res.* **6:** 181.

Elston, R.C., E. Kringlen, and K.K. Namboodiri. 1973. Possible linkage relationships between certain blood groups and schizophrenia or other psychoses. *Behav. Genet.* **3:** 101.

Endicott, J., N. Andreasen, and R.L. Spitzer. 1975. *Family history research diagnostic criteria.* New York State Psychiatric Institute, New York.

Erlenmeyer-Kimling, L., B.A. Cornblatt, D. Rock, S. Roberts, M. Bell, and A. West. 1993. The New York high-risk pro-

ject: Anhedonia, attentional deviance, and psychopathology. *Schizophr. Bull.* **19:** 141.

Fananas, L., P. Moral, M.A. Panadero, and J. Bertranpetit. 1992. Complement genetic markers in schizophrenia: C3, BF and C6 polymorphisms. *Hum. Hered.* **42:** 162.

Fananas, L., P. Moral, J. Tusquets, L. Marti, and J. Bertranpetit. 1990. Genetic markers in schizophrenia: ACPI, ESD, TF and GC polymorphisms. *Hum. Hered.* **40:** 136.

Fang, N., H. Coon, M. Hoff, M., J. Holik, D. Hadley, F. Reimherr, P. Wender, M. Myles-Worsley, M. Waldo, R. Freedman, and W. Byerley. 1995. Search for a schizophrenia susceptibility gene on chromosome 18. *Psychiatr. Genet.* **5:** 31.

Feighner, J.P., E. Robins, S.B. Guze, R.A. Woodruff, G. Winokur, and R. Munoz. 1972. Diagnostic criteria for use in psychiatric research. *Arch. Gen. Psychiatry* **26:** 57.

Fountain, J.W., M.R. Wallace, M.A. Bruce, B.R. Seizinger, A.G. Menon, J.F.Gusella, V.V. Michels, M.A. Schmidt, G.W. Dewald, and F.S. Collins. 1989. Physical mapping of a translocation breakpoint in neurofibromatosis. *Science* **244:**1085.

Garofalo, G., R.M. Ragusa, A. Argiolas, C. Scavuzzo, E. Spina, and C. Barletta. 1993. Evidence of chromosomal fragile sites in schizophrenic patients. *Ann. Genet.* **36:** 132.

Garver, D.L., T. Reich, K.E. Isenberg, and C.R. Cloninger. 1989. Schizophrenia and the question of genetic heterogeneity. *Schizophr. Bull.* **15:** 421.

Ghosh, S. 1995. Probability and complex genes. *Nat. Genet.* **9:** 223.

Gill, M., H. Vallada, C. Collier, P. Sham, P. Holmans, R. Murray, P. McGuffin, S. Nanko, M. Owen, S. Antonarakis, D. Housman, H. Kazazian, G. Nestadt, A.E. Pulver, R.E. Straub, C.J. MacLean, D. Walsh, K.S. Kendler, L. Delisi, M. Polymeropoulos, H. Coon, W. Byerley, R. Lofthouse, E. Gershon, L. Golden, and more than 22 others. 1996. A combined analysis of D22S278 marker alleles in affected sib-pairs support for a susceptibility locus for schizophrenia at chromosome 22q12. *Am. J. Med. Genet.* **67:** 40.

Goldstein, J.M., M.T. Tsuang, and S.V. Faraone. 1989. Gender and schizophrenia: Implications for understanding the heterogeneity of the illness. *Psychiatry Res.* **28:** 243.

Gordon, C.T., D. Krasnewich, B. White, M. Lenane, and J.L. Rapoport. 1994. Brief report: Translocation involving chromosomes 1 and 7 in a boy with childhood-onset schizophrenia. *J. Autism Dev. Disorders* **24:** 537.

Gottesman, I.I. and J. Shields. 1982. *The epigenetic puzzle.* Cambridge University Press, New York.

Green, M.F., H.S. Bracha, P. Satz, and C.D. Christenson. 1994. Preliminary evidence for an association between minor physical anomalies and second trimester neurodevelopment in schizophrenia. *Psychiatry Res.* **53:** 119.

Griffon, N., M.A. Crocq, C. Pilon, M.P. Martres, A. Mayerova, G. Uyanik, E. Burgert, F. Duval, J.P. Macher, F. Javoy-Agid, C.A. Tamminga, J.C. Schwartz, and P. Sokoloff. 1996. Dopamine D3 receptor gene: Organization transcript variants, and polymorphism associated with schizophrenia. *Am. J. Med. Genet.* **67:** 63.

Gur, R.E. and G.D. Pearlson. 1993. Brain imaging in schizophrenia. A review. *Schizophr. Bull.* **19:** 337.

Gurling, H., G. Kalsi, A. Hui-Sui Chen, M. Green, R. Butler, T. Read, P. Murphy, D. Curtis, T. Sharma, H. Petursson, and J. Brynjolfsson. 1995. Schizophrenia susceptibility and chromosome 6p24-22. *Nat. Genet.* **11:** 234.

Gusella, J.F., N.S. Wexler, P.M. Conneally, S.L. Naylor, M.A. Anderson, R.E. Tanzi, P.C. Watkins, K. Ottina, M.R. Wallace, A.Y. Sakaguchi, A.B. Young, I. Shoulson, E. Bonilla, and J.B. Martin. 1983. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306:** 234.

Hattori, M. and S. Nanko. 1995. Association of neurotrophin-3 gene variant with severe forms of schizophrenia. *Biochem. Biophys. Res. Commun.* **209:** 513.

Holland, T. and C. Gosden. 1990. A balanced chromosomal translocation partially co-segregating with psychotic illness in a family. *Psychiatry Res.* **32:** 1.

Inayama, Y., H. Yoneda, T. Sakai, T. Ishida, Y. Nonomura, Y. Kono, R. Takahata, J. Koh, J. Sakai, A. Takai, Y. Inada, and H. Asaba. 1996. Positive association between a DNA sequence variant in the serotonin 2A receptor gene and schizophrenia. *Am. J. Med. Genet.* **67:** 103.

Itokawa, M., T. Arinami, N. Futamura, H. Hamaguchi, and M. Toru. 1993. A structural polymorphism of human dopamine D2 receptor, D2(Ser311→Cys). *Biochem. Biophys. Res. Commun.* **196:** 1369.

Jakob, H. and H. Beckmann 1986. Prenatal developmental disturbances in the limbic allocortex in schizophrenics. *J. Neural Transm.* **65:** 303.

Jones, P. and R.M. Murray. 1991. The genetics of schizophrenia is the genetics of neurodevelopment. *Br. J. Psychiatry* **158:** 615.

Jones, P., B. Rodgers, R. Murray, and M. Marmot. 1994. Child development risk factors for adult schizophrenia in the British 1946 birth cohort. *Lancet* **344:** 1398.

Kalsi, G., D. Curtis, J. Brynjolfsson, R. Butler, T. Sharma, P. Murphy, T. Read, H. Petursson, and H.M. Gurling. 1995. Investigation by linkage analysis of the XY pseudoautosomal region in the genetic susceptibility to schizophrenia. *Br. J. Psychiatry* **167:** 390.

Karayiorgou, M., M.A. Morris, B. Morrow, R.J. Shprintzen, R. Goldberg, J. Borrow, A. Gos, G. Nestadt, P.S. Wolyniec, V.K. Lasseter, H. Eisen, B. Childs, H.H. Kazazian, R. Kucherlapati, S.E. Antonarakis, A.E. Pulver, and D.E. Housman. 1995. Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proc. Natl. Acad. Sci.* **92:** 7612.

Katsanis, J., J. Ficken, W.G. Iacono, and M. Beiser. 1992. Season of birth and electrodermal activity in functional psychoses. *Biol. Psychiatry* **31:** 841.

Kendler, K.S. and S. Diehl. 1993. The genetics of schizophrenia: A current, genetic-epidemiologic perspective. *Schizophr. Bull.* **19:** 261.

Kendler, K.S., C.C. Masterson, R. Ungaro, and K.L. Davis. 1984. A family history study of schizophrenia-related personality disorders. *Am. J. Psychiatry* **41:** 424.

Kendler, K.S., M. McGuire, A. M. Gruenberg, A. O'Hare, M. Spellman, and D. Walsh. 1993. The Roscommon family study. III. Schizophrenia-related personality disorders in relatives. *Arch. Gen. Psychiatry* **50:** 781.

Kennedy, J.L., E.A. Billett, F.M. Macciardi, M. Verga, T.J. Parsons, H.Y. Meltzer, J. Lieberman, and J.A. Buchanan. 1995. Association study of dopamine d3 receptor gene and schizophrenia. *Am. J. Med. Genet.* **60:** 558.

Kennedy, J.L., L.A. Giuffra, H.W. Moises, L.L. Cavalli-Sforza, A.J. Pakstis, J.R. Kidd, C.M. Castiglione, B. Sjogren, L. Wetterberg, and K.K. Kidd. 1988. Evidence against linkage of schizophrenia to markers on chromosome 5 in a northern Swedish pedigree. *Nature* **336:** 166.

Keshavan, M.S., S. Anderson, and J.W. Pettegrew. 1994. Is schizophrenia due to excessive synaptic pruning in the prefrontal cortex? The Feinberg hypothesis revisited. *J.*

Psychiatr. Res. **28:** 239.

Kringlen, E. and G. Cramer. 1989. Offspring of monozygotic twins discordant for schizophrenia. *Arch. Gen. Psychiatry* **46:** 873.

Kruglyak, L. and E.S. Lander. 1995. Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am. J. Hum. Genet.* **57:** 439.

Kruglyak, L., M.J. Daly, M.P. Reeve-Daly, and E.S. Lander. 1996. Parametric and nonparametric linkage analysis: A unified multipoint approach. *Am. J. Hum. Genet.* **58:** 1347.

Kuokkanen, S., M. Sundvall, J.D. Terwilliger, P.J. Tienari, J. Wikstrom, R. Holmdahl, U. Pettersson, and L. Peltonen. 1996. A putative vulnerability locus to multiple sclerosis maps to 5p14-p12 in a region syntenic to the murine locus Eae2. *Nat. Genet.* **13:** 477.

Lasseter, V.K., A.E. Pulver, P.S. Wolyniec, G. Nestadt, D. Meyers, M. Karayiorgou, D. Housman, S. Antonarakis, H. Kazazian, L. Kasch, R. Babb, M. Kimberland, and B. Childs. 1995. Follow-up report of potential linkage for schizophrenia on chromosome 22q (part 3). *Am. J. Med. Genet.* **60:**172.

Lathrop, G.M. and J.M. Lalouel. 1984. Easy calculations of lod scores and genetic risks on small computers. *Am. J. Hum. Genet.* **36:** 460.

Lathrop, G.M., J.M. Lalouel, and R.L. White. 1986. Construction of human linkage maps: Likelihood calculations for multilocus linkage analysis. *Genet. Epidemiol.* **3:** 39.

Lathrop, G.M., J.M. Lalouel, C. Julier, and J. Ott. 1984. Strategies for multilocus linkage analysis in humans. *Proc. Natl. Acad. Sci.* **81:** 3443.

Levinson, D.F., D.B. Wildenauer, S.G. Schwab, M. Albus, J. Hallmayer, B. Lerer, W. Maier, D. Blackwood, W. Muir, D. St. Clair, S. Morris, H.W. Moises, L. Yang, H. Kristbjarnarson, T. Helgason, C. Wiese, D.A. Collier, P. Holmans, J. Daniels, M. Rees, P. Asherson, Q. Roberts, A. Cardno, M.J. Arranz, H. Vallada, et al. 1996. Additional support for schizophrenia linkage on chromosomes 6 and 8—A multicenter study. *Am. J. Med. Genet.* **67:** 580.

Lewis, S. 1989. Congenital risk factors for schizophrenia. *Psychol. Med.* **19:** 5.

Liang, K.Y. and A.E. Pulver. 1996. Analysis of case-control/family sampling design. *Genet. Epidemiol.* **13:** 253.

Lin, M.W., D. Curtis, N. Williams, M. Arranz, S. Nanko, D. Collier, P. McGuffin, R. Murray, M. Owen, M. Gill, and J. Powell. 1995. Suggestive evidence for linkage of schizophrenia to markers on chromosome 13q14.1-q32. *Psychr. Genet.* **5:** 117.

Lindsay, E.A., P. Rizzu, R. Antonacci, V. Jurecic, J. Delmas-Mata, C.C. Lee, U.J. Kim, P.J. Scambler, and A. Baldini. 1996. A transcription map in the CATCH22 critical region: Identification, mapping, and ordering of four novel transcripts expressed in heart. *Genomics* **32:** 104.

Lindsay, E.A., M.A. Morris, A. Gos, G. Nestadt, P.S. Wolyniec, V.K. Lasseter, R. Shprintzen, S.E. Antonarakis, A. Baldini, and A.E. Pulver. 1995. Schizophrenia and chromosomal deletions within 22q11.2. *Am. J. Hum. Genet.* **56:** 1502.

Lyon, M., C.E. Barr, T.D. Cannon, S.A. Mednick, and D. Shore. 1989. Fetal neural development and schizophrenia. *Schizophr. Bull.* **15:** 149.

Macciardi, F., J.L. Kennedy, L. Ruocco, L. Giuffra, P. Carrera, C. Marino, V. Rinaldi, E. Smeraldi, and M. Ferrari. 1992. A genetic linkage study of schizophrenia to chromosome 5 markers in a northern Italian population. *Biol. Psychiatry* **31:** 720.

Maier, W., F. Schmidt, S.G. Schwab, J. Hallmayer, J. Minges, M. Ackenheil, D. Lichtermann, and D.B. Wildenauer. 1995. Lack of linkage between schizophrenia and markers at the telomeric end of the pseudoautosomal region of the sex chromosomes. *Biol. Psychiatry* **37:** 344.

Malaspina, D., D. Warburton, X. Amador, M. Harris, and C.A. Kaufmann. 1992. Association of schizophrenia and partial trisomy of chromosome 5p. A case report. *Schizophr. Res.* **7:** 191.

Marsh, L., R.L. Suddath, N. Higgins, and D.R. Weinberger. 1994. Medial temporal lobe structures in schizophrenia: Relationship of size to duration of illness. *Schizophr. Res.* **11:** 225.

Maziade, M., M. Debraekeleer, P. Genest, D. Cliche, J.P. Fournier, Y. Garneau, C. Shriqui, M.A. Roy, L. Nicole, V. Raymond, and M. Vekemans. 1993. A balanced 2:18 translocation and familial schizophrenia: Falling short of an association. *Arch. Gen. Psychiatry* **50:** 73.

McGue, M. and I.I. Gottesman. 1991. The genetic epidemiology of schizophrenia and the design of linkage studies. *Eur. Arch. Psychiatry Clin. Neurosci.* **240:** 174.

McGuffin, P., M. Sargeant, G. Hetti, S. Tidmarsh, S. Whatley, and R.M. Marchbanks. 1990. Exclusion of a schizophrenia susceptibility gene from the chromosome 5q11-q13 region: New data and a reanalysis of previous reports. *Am. J. Hum. Genet.* **47:**524.

McNeil, T.F. 1995. Perinatal risk factors and schizophrenia: selective review and methodological concerns. *Epidemiol. Rev.* **17:** 107.

Moises, H.W., L. Yang, T. Li, B. Havsteen, R. Fimmers, M.P. Baur, X.H. Liu, and I.I. Gottesman. 1995a. Potential linkage disequilibrium between schizophrenia and locus D22S278 on the long arm of chromosome 22. *Am. J. Med. Genet.* **60:** 465.

Moises, H.W., L. Yang, H. Kristbjarnarson, C. Wiese, W. Byerley, F. Macciardi, V. Arolt, D. Blackwood, X. Liu, B. Sjogren, H.N. Auchauer, H.G. Hwu, K. Jang, W.J. Livesley, J.L. Kennedy, T. Zoega, O. Ivarsson, M.T. Bui, M.H. Yu, B. Havsteen, D. Commenges, J. Weissenbach, E. Schwinger, I.I. Gottesman, A.J. Pakstis, L. Wetterberg, K.K. Kidd, and T. Helgason. 1995b. An international two-stage genome-wide search for schizophrenia susceptibility genes. *Nat. Genet.* **11:** 321.

Murray, R.M. 1994. Neurodevelopmental schizophrenia: The rediscovery of dementia praecox. *Br. J. Psychiatry* **25:** 6.

Murray, R.M. and S.W. Lewis. 1987. Is schizophrenia a neurodevelopmental disorder? *Br. Med. J.* **285:** 681.

Murray, R.M. and E. O'Callaghan. 1991. Neurodevelopmental schizophrenia. *Schizophr. Monitor* **1:** 1.

Murray, R.M., A.M. Revellry, and P. McGuffin. 1986. Genetic vulnerability to schizophrenia. *Schizophr. Bull.* **9:** 3.

Nanko, S., H. Kunugi, T. Sasaki, R. Fukuda, T. Kawate, and H. Kazamatsuri. 1993. Pericentric region of chromosome 9 is a possible candidate region for linkage study of schizophrenia. *Biol. Psychiatry* **33:** 655.

Nanko, S., M. Hattori, S. Kuwata, T. Sasaki, R. Fukuda, X.Y. Dai, K. Yamaguchi, Y. Shibata, and H. Kazamatsuri. 1994. Neurotrophin-3 gene polymorphism associated with schizophrenia. *Acta Psychiatr. Scand.* **89:** 390.

Nimgaonkar, V.L., X.R. Zhang, J.G. Caldwell, R. Ganguli, and A. Chakravarti. 1993a. Association study of schizophrenia with dopamine D3 receptor gene polymorphisms: Probable effects of family history of schizophrenia? *Am. J. Med. Genet.* **48:** 214.

Nimgaonkar, V.L., R. Ganguli, W.A. Rudert, C. Vavassori,

B.S. Rabin, and M. Trucco. 1993b. A negative association of schizophrenia with an allele of the HLA DQB1 gene among African-Americans. *Schizophr. Res.* **8:** 199.

Nimgaonkar, V.L., W.A. Rudert, X.R. Zhang, W.F. Tsoi, M. Trucco, and N. Saha. 1995. Further evidence for an association between schizophrenia and the HLA DQB1 gene locus. *Schizophr. Res.* **18:** 43.

Okoro, C., R. Bell, P. Sham, S. Nanko, P. Asherson, M. Owen, M. Gill, P. McGuffin, R.M. Murray, and D. Collier. 1995. No evidence for linkage between the X-chromosome marker DXS7 and schizophrenia. *Am. J. Med. Genet.* **60:** 461.

Ott, J. 1991. *Analysis of human genetic linkage*, revised edition. The Johns Hopkins University Press, Baltimore.

Owen, M.J. 1992. Will schizophrenia become a graveyard for molecular geneticists? *Psychol. Med.* **22:** 289.

Palmour, R.M., S. Miller, A. Fielding, M. Vekemans, and F.R. Ervin. 1994. A contribution to the differential diagnosis of the "group of schizophrenias," Structural abnormality of chromosome 4. *J. Psychiatry Neruosci.* **19:** 270.

Parnas, J. 1985. Mates of schizophrenic mothers: A study of assortative mating from the American-Danish high-risk project. *Br. J. Psychiatry* **140:** 490.

———. 1988. Assortative mating in schizophrenia: Results from the Copenhagen high-risk study. *Psychiatry* **51:** 58.

Petty, R.G., P.E. Barta, G.D. Pearlson, I.K. McGilchrist, R.W. Lewis, A.Y. Tien, A. Pulver, D.D. Vaughn, M.F. Casanova, and R.E. Powers. 1995. Reversal of asymmetry of the planum temporale in schizophrenia. *Am. J. Psychiatry* **152:** 715.

Polymeropoulos, M.H., H. Coon, W. Byerley, E.S. Gershon, L. Goldin, T.J. Crow, J. Rubenstein, M. Hoff, J. Holik, A.M. Smith, G. Sheilds, N.J. Bass, M. Poulter, R. Lofthouse, A. Vita, C. Morganti, C.R. Merril, and L.E. DeLisi. 1994. Search for a schizophrenia susceptibility locus on human chromosome 22. *Am. J. Med. Genet.* **54:** 93.

Pulver, A.E. and S.J. Bale. 1989. Availability of schizophrenic patients and their families for genetic linkage studies: Findings from the Maryland epidemiology sample. *Genet. Epidemiol.* **6:** 671.

Pulver, A.E., K.Y. Liang, P.S. Wolyniec, J. McGrath, B.A. Melton, L. Adler, and B. Childs. 1992a. Season of birth of siblings of schizophrenic patients (comments). *Br. J. Psychiatry* **160:** 71.

Pulver, A.E., C.H. Brown, P. Wolyniec, J. McGrath, D. Tam, L. Adler, W.T. Carpenter, and B. Childs. 1990. Schizophrenia: Age at onset, gender and familial risk. *Acta Psychiatr. Scand.* **82:** 344.

Pulver, A.E., K.Y. Liang, C.H. Brown, P. Wolyniec, J. McGrath, L. Adler, D. Tam, W.T. Carpenter, and B. Childs. 1992b. Risk factors in schizophrenia season birth, gender, and familial risk. *Br. J. Psychiatry* **160:** 65.

Pulver, A.E., G. Nestadt, R. Goldberg, R.J. Shprintzen, M. Lamacz, P.S. Wolyniec, B. Morrow, M. Karayiorgou, S.E. Antonarakis, D. Housman, and R. Kucherlapati. 1994a. Psychotic illness in patients diagnosed with velo-cardio-facial syndrome and their relatives. *J. Nerv. Ment. Dis.* **182:** 476.

Pulver, A.E., V.K. Lasseter, L. Kasch, P. Wolyniec, G. Nestadt, J.L. Blouin, M. Kimberland, R. Babb, S. Vourlis, H. Chen, M. Lalioti, M.A. Morris, M. Karayiorgou, J. Ott, D. Meyers, S.E. Antonarakis, D. Housman, and H.H. Kazazian. 1995. Schizophrenia: A genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes. *Am. J. Med. Genet.* **60:** 252.

Pulver, A.E., M. Karayiorgou, P.S. Wolyniec, V.K. Lasseter, L. Kasch, G. Nestadt, S. Antonarakis, D. Housman, H.H. Kazazian, D. Meyers, J. Ott, M. Lamacz, K.Y. Liang, J. Hanfelt, G. Ullrich, N. DeMarchi, E. Ramu, P.R. McHugh, L. Adler, M. Thomas, W.T. Carpenter, T. Manschreck, C.T. Gordon, M. Kimberland, R. Babb, J. Puck, and B. Childs. 1994b. Sequential strategy to identify a susceptibility gene for schizophrenia: Report of potential linkage on chromosome 22q12-q13.1: Part 1. *Am. J. Med. Genet.* **54:** 36.

Pulver, A.E., M. Karayiorgou, V.K. Lasseter, P. Wolyniec, L. Kasch, S. Antonarakis, D. Housman, H.H. Kazazian, D. Meyers, G. Nestadt, J. Ott, K.Y. Liang, M. Lamacz, M. Thomas, B. Childs, S.R. Diehl, S. Wang, B. Murphy, C. Sun, F.A. O'Neill, L. Nie, P. Sham, J. Burke, B.W. Duke, F. Duke, and 29 others. 1994c. Follow-up of a report of a potential linkage for schizophrenia on chromosome 22q12-q13.1 (part 2). *Am. J. Med. Genet.* **54:** 44.

Rao, D.C., N.E. Morton, I.I. Gottesman, and R. Lew. 1981. Path analysis of qualitative data on pairs of relatives: Application to schizophrenia. *Hum. Hered.* **31:** 325.

Rimmer, J. and B. Jacobsen. 1976. Differential fertility of adopted schizophrenics and their half-siblings. *Acta Psychiatr. Scand.* **54:** 161.

Rios, R., F. Sole, C. Montes, A. Vicente, M.M. Perez, M. Valle, and F. Gascon. 1995. A new case of trisomy 5 as sole cytogenetic anomaly in acute myleoid leukemia. *Cancer Genet. Cytogenet.* **84:** 120.

Risch, N. and M. Baron. 1984. Segregation analysis of schizophrenia and related disorders. *Am. J. Hum. Genet.* **36:** 1039.

Roberts, G.W. 1991. Schizophrenia: A neuropathological perspective. *Br. J. Psychiatry* **158:** 8.

Ross, C.A. and G.D. Pearlson. 1996. Schizophrenia, the heteromodal association neocortex and development: Potential for a neurogenetic approach. *Trends Neurosci.* **19:** 171.

Sacchetti, E., A. Calzeroni, A. Vita, A. Terzi, F. Pollastro, and C.L. Cazzullo. 1992. The brain damage hypothesis of the seasonality of births in schizophrenia and major affective disorders: Evidence from computerised tomography. *Br. J. Psychiatry* **160:** 390.

Sacker, A., D.J. Done, T.J. Crow, and J. Golding. 1995. Antecedents of schizophrenia and affective illness. Obstetric complications (comments). *Br. J. Psychiatry* **166:** 734.

Saha, N., J.S.H. Tay, W.F. Tsoi, and E.H. Kua. 1990. Association of Duffy blood group with schizophrenia in Chinese (letter). *Genet. Epidemiol.* **7:** 303.

Sanders, A.R., D.E. Rincon-Limas, R. Chakraborty, B. Grandchamp, J.D. Hamilton, W.E. Fann, and P.I. Patel. 1992. Association between genetic variation at the porphobilinogen deaminase gene and schizophrenia. *Schizophr. Res.* **8:** 211.

Satsangi, J., M. Parkes, E. Louis, L. Hashimoto, N. Kato, K. Welsh, J.D. Terwilliger, G.M. Lathrop, J.I. Bell, and D.P. Jewell. 1996. Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7, and 12. *Nat. Genet.* **14:** 199.

Schaffer, A.A., S.K. Gupta, K. Shriram, and R.W. Cottingham. 1994. Avoiding recomputation in linkage analysis. *Hum. Hered.* **44:** 225.

Schwab, S.G., M. Albus, J. Hallmayer, S. Honig, M.

Borrmann, D. Lichtermann, R.P. Ebstein, M. Ackenheil, B. Lerer, N. Risch, W. Maier, and D.B. Wildenauer. 1995a. Evaluation of a susceptibility gene for schizophrenia on chromosome 6p by multipoint affected sib-pair linkage analysis. *Nat. Genet.* **11:** 325.

Schwab, S.G., B. Lerer, M. Albus, W. Maier, J. Hallmayer, R. Fimmers, D. Lichtermann, J. Minges, B. Bondy, M. Ackenheil, D. Altmark, D. Hasib, E. Gur, R.P. Ebstein, and D.B. Wildenauer. 1995b. Potential linkage for schizophrenia on chromosome 22Q12- Q13—A replication study. *Am. J. Med. Genet.* **60:** 436.

Sherrington, R., J. Brynjolfsson, H. Petursson, M. Potter, K. Dudleston, B. Barraclough, J. Wasmuth, M. Dobbs, and H. Gurling. 1988. Localization of a susceptibility locus for schizophrenia on chromosome 5. *Nature* **366:** 164.

Spielman, R.S., R.E. McGinnis, and W.J. Ewens. 1993. Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* **52:** 506.

St. Clair, D., D. Blackwood, W. Muir, D. Baillie, A. Hubbard, A. Wright, and H.J. Evans. 1989. No linkage of chromosome 5q11-q13 markers to schizophrenia in Scottish families. *Nature* **339:** 305.

Straub, R.E., C.J. MacLean, F.A. O'Neill, J. Burke, B. Murphy, F. Duke, R. Shinkwin, B.T. Webb, J. Zhang, D. Walsh, and K.S. Kendler. 1995. A potential vulnerability locus for schizophrenia on chromosome 6p24-22: Evidence for genetic heterogeneity. *Nat. Genet.* **11:** 287.

Susser, E., R. Neugebauer, H.W. Hoek, A.S. Brown, S. Lin, D. Labovitz, and J.M. Gorman. 1996. Schizophrenia after prenatal famine. Further evidence. *Arch. Gen. Psychiatry* **53:** 25.

Taylor, M.A. 1992. Are schizophrenia and affective disorder related? A selective literature review. *Am. J. Psychiatry* **149:** 22.

Terwilliger, J.D. 1994.The available possibilities to analyze data of polygenic disease statistically. In *IV Workshop of the Nordic Genome Initiative,* August (Abstr.).

Tsuang, M.T. and S.V. Faraone. 1995. The case for heterogeneity in the etiology of schizophrenia. *Schizophr. Res.* **17:** 161.

Tsuang, M.T., M.W. Gilbertson, and S.V. Faraone. 1991. The genetics of schizophrenia current knowledge and future directions. *Schizophr. Res.* **4:** 157.

Wang, S., C.E. Sun, C.A. Walczak, J.S. Ziegle, B.R. Kipps, L.R. Goldin, and S.R. Diehl. 1995. Evidence for a susceptibility locus for schizophrenia on chromosome 6pter-p22. *Nat. Genet.* **10:** 41.

Wang, Z.W., D. Black, N. Andreasen, and R.R. Crowe. 1993. Pseudoautosomal locus for schizophrenia excluded in 12 pedigrees. *Arch. Gen. Psychiatry* **50:** 199.

Wei, J., C.N. Ramchand, and G.P. Hemmings. 1995. Association of polymorphic VNTR region in the first intron of the human TH gene with disturbances of the catecholamine pathway in schizophrenia. *Psychiatr. Genet.* **5:** 83.

Weinberger, D.R. 1987. Implications of normal brain development for the pathogenesis of schizophrenia. *Arch. Gen. Psychiatry* **44:** 660.

Wolyniec, P.S., A.E. Pulver, J.A. McGrath, and D. Tam. 1992. Schizophrenia: Gender and familial risk. *J. Psychiatr. Res.* **26:** 17.

Zamani, M.G., M. De Hert, M. Spaepen, M. Hermans, P. Marynen, J.J. Cassiman, and J. Peuskens. 1994. Study of the possible association of HLA class II, CD4, and CD3 polymorphisms with schizophrenia. *Am. J. Med. Genet.* **54:** 372.

Exhibit L

It is now well established that genetic factors play an important role in the etiology of psychiatric diseases such as schizophrenia and bipolar disorder, although their etiology is not yet known. In order to further understand these diseases, from 1980 to 1992, the European Science Foundation helped to fund a large consortium of 22 institutions from 14 European countries which brought together pedigrees from over 150 families suffering from one of these two diseases. So that the study would be uniform, all participating groups used the same diagnostic instruments (SCAN/PSE 10 and the Schedule for Affective Disorders and Schizophrenia-Lifetime) which in turn were centralized using a computerized checklist, OPCRIT.

Genotypings using 400 markers for the entire genome and for almost 1,600 samples were carried out by the Genethon laboratory in Evry, France. The statistical analyses based on these 640,000 genotypings using lod score and sib pair methods are currently in progress and will be discussed.

PRELIMINARY RESULTS OF AN INTERNATIONAL GENOMIC SCAN FOR SCHIZOPHRENIA. L.E. DeLisi, M. Kelly, S. Shaw, L. Cardon, A.B. Smith, P. Hopkins, G. Shields, J. Loftus, S. Laval, T.J. Crow, and R. Sherrington. Department of Psychiatry, SUNY, Stony Brook, NY 11733; Sequana Therapeutics, Inc., La Jolla, California; and the Department of Psychiatry, Oxford, United Kingdom.

An initial cohort of 80 nuclear families, which consisted of 131 sib pairs with DSM-III-R schizophrenia or schizoaffective disorder, had DNA examined for linkages to 341 polymorphic PCR markers at Sequana Therapeutics, Inc., using a systematic genomic screening approach. Markers were spaced approximately 10 cM apart, with no gap greater than 25 cM. Description of the internationally based clinical sample used and laboratory procedures have been previously reported [Garner et al., Am J Med Genet 1996;67:595–610].

Two-point nonparametric preliminary analyses were performed using the program, Genehunter. Three separate sets of results were examined for each of 3 models for illnesses: 1) chronic schizophrenia only (N = 48 families), 2) chronic schizophrenia or schizoaffective disorder (N = 32 families), and 3) a broad category extending the diagnosis to the above plus paranoid or schizotypal personality disorders and other nonaffective psychoses within the 80 families.

Using a significance level of 0.05 to identify loci of potential interest for further investigations, 15 were found on 10 separate chromosomes. The 3 most significant regions were 1) 4q21.1-q22 with an NPL of 2.33 with marker D4S1546 and a broad Dx ($P < .009$), 2) 22q11-q11.2 with an NPL of 2.16 with marker D22S446 and either a broad or schizophrenia/schizoaffective Dx ($P < .01$), and 3) 2p14-p15 with an NPL of 2.02 and a broad Dx ($P < .02$).

Complete parametric analyses for all marker data are now in progress. All regions with statistical significance at the $P = .01$ level will be examined further with more densely placed markers in the larger cohort of 300 families and a new Costa Rican isolated cohort of large sibships.

GENOME SCAN FOR SCHIZOPHRENIA GENES: A DETAILED PROGRESS REPORT IN AN IRISH COHORT. R.E. Straub,[1,3] C.J. MacLean,[1,2,3] F.A. O'Neill,[4] D. Walsh,[5] and K.S. Kendler.[1,2,3] [1]Department of Psychiatry, Virginia Commonwealth University, Richmond, VA; [2]Department of Human Genetics, Virginia Commonwealth University, Richmond, VA; [3]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA; [4]The Department of Psychiatry, The Queens University, Belfast, Northern Ireland; [5]The Health Research Board, Dublin, Ireland.

The Irish Study of High Density Schizophrenia Families (ISHDSF) contains 1,408 individuals with DNA in 265 families systematically ascertained from psychiatric facilities in Ireland and Northern Ireland. Using linkage analysis, we have found support for the presence of loci influencing the susceptibility to schizophrenia in three chromosomal regions: 6p24-21, 8p22-21, and 5q21-31. In order to identify regions containing additional susceptibility genes, we are currently doing both a genome scan and follow-up genotyping. For the scan, we have divided the families randomly into three sets, and are genotyping different family sets with adjacent markers. We are using markers at approximately 20 cM intervals at first, and then adding intedigitating markers to create a final resolution of <10 cM. For follow-up, we have been more carefully examining many potentially positive regions—those indicated by our own scan and those suggested by other investigators. This is done by testing more markers and by genotyping additional sets of families. The final marker map will, therefore, in many regions by much more dense than the overall maximum final intermarker distance of 10 cM. Detailed results from both parametric and nonparametric linkage analysis will be provided.

SEARCHING FOR SUSCEPTIBILITY GENES IN SCHIZOPHRENIA BY AFFECTED SIB-PAIR ANALYSIS (GERMANY). D.B. Wildenauer,[1] M. Albus,[2] S.G. Schwab,[1] J. Hallmayer,[3] C. Hanses,[1] G.N. Eckstein,[1] P. Zill,[4] S. Hönig,[4] B. Lerer,[5] R. Ebstein,[6] D. Lichtermann,[1] M. Trixler,[1] M. Borrmann,[2] and W. Maier.[1] [1]Department of Psychiatry, University of Bonn, Bonn, Germany; [2]Mental State Hospital, Haar, Germany; [3]Department of Psychiatry, University of Perth, Perth, Australia; [4]Department of Psychiatry, University of Munich, Munich, Germany; [5]Hadassah Medical School, Jerusalem, Israel; [6]Sarah Herzog Hospital, Jerusalem, Israel; [7]Department of Psychiatry, University of Pecs, Pecs, Hungary.

Our sample comprises 72 families with the index patient suffering from schizophrenia or chronic schizoaffective disorder (schizophrenic type). The minimum requirement for including a family in the sample has been two affected siblings with parental information for analysis identity by descent. Thirteen families had been ascertained in Israel (mainly Sephardic origin), 2 in Hungary, and 57 in Germany. A genome scan has been initiated using fluorescence technology and an ABI 377 for detection and genotyping. For first analysis, the average distance between markers was 10–20 cM. Areas of interest were rescanned with markers in a distance of 2 cM and less. To date, we have finished the scan for chromosomes 1,2,3,4,5,6,11,13,15,18,19,21, and 22. Four potential susceptibility loci have been detected in our family sample mapping to 5q31, 6p22,23, 18p, and 22q12,13.

Supported by Deutsche Forschungsgemeinschaft.

Exhibit M

American Journal of Medical Genetics (Neuropsychiatric Genetics) 81:364–376 (1998)

A Genome-Wide Search for Schizophrenia Susceptibility Genes

Sarah H. Shaw,[1] Mary Kelly,[1] Angela B. Smith,[2] Gail Shields,[2] Penelope J. Hopkins,[1] Josephine Loftus,[3] Steven H. Laval,[3] Antonio Vita,[4] Marc De Hert,[5] Lon R. Cardon,[1] Timothy J. Crow,[3] Robin Sherrington,[1] and Lynn E. DeLisi[2]*

[1]*Axys Pharmaceuticals, La Jolla, California*
[2]*Department of Psychiatry, SUNY, Stony Brook, New York*
[3]*University Department of Psychiatry, Warneford Hospital, Oxford, UK*
[4]*Department of Psychiatry, University of Milan, Milan, Italy*
[5]*University Hospital St. Jozef, Leuvensesteenweg, Belgium*

We completed a systematic genome-wide search for evidence of loci linked to schizophrenia using a collection of 70 pedigrees containing multiple affected individuals according to three phenotype classifications: schizophrenia only (48 pedigrees; 70 sib-pairs); schizophrenia plus schizoaffective disorder (70 pedigrees; 101 sib-pairs); and a broad category consisting of schizophrenia, schizoaffective disorder, paranoid or schizotypal personality disorder, psychosis not otherwise specified (NOS), delusional disorder, and brief reactive psychosis (70 pedigrees; 111 sib-pairs). All 70 families contained at least one individual affected with chronic schizophrenia according to DSM-III-R criteria. Three hundred and thirty-eight markers spanning the genome were typed in all pedigrees for an average resolution of 10.5 cM (range, 0–31 cM) and an average heterozygosity of 74.3% per marker. The data were analyzed using multipoint nonparametric allele-sharing and traditional two-point lod score analyses using dominant and recessive, affecteds-only models. Twelve chromosomes (1, 2, 4, 5, 8, 10, 11, 12, 13, 14, 16, and 22) had at least one region with a nominal *P* value <0.05, and two of these chromosomes had a nominal *P* value <0.01 (chromosomes 13 and 16), using allele-sharing tests in GENEHUNTER. Five chromosomes (1, 2, 4, 11, and 13) had at least one marker with a lod score >2.0, allowing for heterogeneity. These regions will be saturated with additional markers and investigated in a new, larger set of families to test for replication. Am. J. Med. Genet. (Neuropsychiatr. Genet.) 81:364–376, 1998.

KEY WORDS: linkage; sib-pair analysis; chromosome 2; chromosome 4; chromosome 13; chromosome 22

INTRODUCTION

Schizophrenia is a severely debilitating neuropsychiatric disorder characterized by disturbance of thought, auditory hallucinations, and multiple delusions. Clear genetic susceptibility has been shown by twin concordance, familial risk, and studies of adopted-away offspring of patients with schizophrenia [Kendler, 1988; McGuffin et al., 1984; Kety et al., 1994]. However, no clear mode of inheritance has been established, despite several attempts by segregation analysis using large family data sets. Schizophrenia is therefore assumed to belong to the class of "complex" genetic disorders (such as diabetes), that have a genetic predisposition due to more than one gene that may produce illness independently in different families, or acting together to cause illness in all susceptible individuals. Further complication in the search for genes for this disorder arises from the uncertainty of the diagnostic boundaries determined by the underlying genetic susceptibility. On the basis of family studies [Kendler et al., 1993, 1995], it has been suggested that severe chronic schizophrenia is also genetically related to schizoaffective disorder, other psychoses, and a milder form of symptoms expressed by paranoid and schizotypal personality disorders, but this can only be proven once genetic loci are found.

For the past decade, families with multiple ill members with schizophrenia have been ascertained in different centers around the world, and molecular genetic

Contract grant sponsor: NIMH; Contract grant number: R01444245; Contract grant sponsor: Medical Research Council, UK; Contract grant sponsor: Axys Pharmaceuticals, Inc.

*Correspondence to: Lynn E. DeLisi, M.D., Department of Psychiatry, T-10, HSC, State University of New York at Stony Brook, Stony Brook, NY 11794. E-mail: ldelisi@ccmail.sunysb.edu

Received 25 November 1997; Revised 19 February 1998

linkage studies have been performed. Initially, most groups concentrated on examining a few large families with multiply affected individuals; however, these studies yielded either a few positives that were never independently replicated or completely negative results [e.g., Sherrington et al., 1988; Crowe et al., 1991].

More recently it has been recognized that in the case of complex genetic disorders, nonparametric sib-pair analyses might yield more significant findings [Risch, 1990]. For these analyses, mode of inheritance does not have to be specified, and only clearly ill individuals with no diagnostic uncertainty are used in the analysis.

It has also been recognized that, with the exception of the sex chromosomes [DeLisi and Crow, 1989], there is no a priori hypothesis that would lead one to focus on any given chromosomal region. Thus, the whole genome needs to be systematically screened. A few genome screens of schizophrenia have been published and/or have been presented at meetings thus far and are listed in Table I. Of these, the findings are clearly inconsistent across studies, although some regions appear to have weak positive replications by at least one other independent group. Large collaborative efforts have also been formed in the hope of clarifying some of these weak positives, but the results remain equivocal [Gill et al., 1996; Schizophrenia Collaborative Linkage Group for Chromosomes 3, 6, and 8, 1996].

Our own studies until recently have focused on the X chromosome and continue to show weakly positive lod scores in the pericentromeric region [DeLisi et al., 1994; Dann et al., 1997]. These remain unreplicated. Straub et al. [1995] focused on an initial finding on chromosome 6p. The original finding, while not consistently replicated, has produced a plethora of independent attempts at confirmation, yielding a broad region including a number of weak positives, but also some negative findings, covering at least 40 cM on chromosome 6p [Wang et al., 1995; Antonarakis et al., 1995; Gurling et al., 1995; Schwab et al., 1995; Schizophrenia Linkage Collaborative Group for Chromosomes 3, 6, and 8, 1996; Mowry et al., 1995]. Our own sample of 211 families [Garner et al., 1996] showed no evidence of linkage to markers in this region.

Another region of interest is on chromosome 22q, first suggested by Pulver et al. [1994a]. A region somewhat distal to this has been positive in some studies [Moises et al., 1995b; Coon et al., 1994a; Gill et al., 1996] and not others [Kalsi et al., 1995; Pulver et al., 1994b; Polymeropoulos et al., 1994; Parsian et al., 1997]. Other positive findings have been reported on chromosomes 2, 3, 4p and q, 5p and q, 8p, 10, 13q, and 18p [e.g., Sherrington et al., 1988; Silverman et al., 1996; Kendler et al., 1996; Lin et al., 1995; Antonarakis et al., 1996; Wildenauer, 1996] (see Table I). At present, these are all preliminary findings, and do not reach conventional levels of accepted significance. Some clearly are false positives, while false negatives may have also arisen in studies with not enough sample power to detect linkage.

The following study is a systematic genome-wide screen to determine where and how many suggestive regions of linkage can be found using an initial set of approximately 100 sib-pairs. These results are compared for consistency with others already reported.

MATERIALS AND METHODS

Clinical

Three international sites have been used as bases for the recruitment of families with at least 2 available sibs diagnosed with schizophrenia or schizoaffective disorder. Details of all clinical procedures for this cohort have been previously published [Garner et al., 1996]. Identification of families, and clinical evaluative and diagnostic procedures, were similar in all locations. Recruitment involved several methods: catchment area screening, systematic contact with health professionals at hospital and out-patient facilities within one day's driving time from the center, and advertisement through local and national support orga-

TABLE I. Published and Presented Genome Scans of Schizophrenia as of October 1997*

References	No. of families	No. affected (SZ or SA)	No. of markers	Spacing	Chromosomal regions with lod scores >2 or $P < 0.01$
Pulver et al. [1995] Antonarakis et al. [1996]	57	161	520	20 cM (70%)	3p26–p24, 8p22–p21, 13q32, 22q11–q12
Moises et al. [1995a]	I: 5 II: 65	37 213	413	9 cM (84%)	2p25–21, 4q12–13, 6p21, 9p23–p21 20p12
Barr et al. [1994]	7	31	180	10–20 cM	None
Coon et al. [1994b]	9	35	329	Unlisted	4p11–q11, 22q11–qter
Levinson et al. [1996]	43	107	335	10 cM	2q, 10q
NIMH Genetics Initiative [1997]	30[a]	79	459	8–10 cM	8p, 2q, 10p
	43[b]	96			
Williams et al. [1997]	100	200	220	20 cM	10q
Straub et al. [1997a]	265	577	Unlisted	<10 cM	5q21–q31, 6p24–p21, 8p22–p21, 10p
Wildenauer et al. [1997]	59	155	Unlisted	10–20 cM	5q31, 6p22–p23, 18p
Byerley et al. [1997]	17	79	391	10 cM	2p13–p14
Horvatta et al. [1997]	1	17	352	10 cM	1, 4
Present study and DeLisi et al. [1997]	70	171	338	10.5 cM	1q22–q23, 2p14–p13, 4p, 5p14, 10q22–q24, 11p, 13q12–q13, 16q22, 22q11

*%, amount of genome covered. Findings are present on chromosomes 1q, 2p, 2q, 3p, 4p, 4q, 5p, 5q, 6p, 8p, 9p, 10p, 10q, 11p, 13q, 16q, 20p, and 22q.
[a]African-American.
[b]European-American.

nizations for families of the mentally ill (i.e., the National Alliance for Mental Illness (NAMI) in the USA, and the Schizophrenia A National Emergency (SANE) in the UK). For the present genome scan and laboratory analysis, a total of 70 families was examined: 42 families from the USA, 14 from the UK and Ireland, 12 from Northern Italy, and 2 from Belgium. Only families of European descent and Caucasian race were included, and of these, only those families in which at least one member of an ill sib-pair had DSM-III-R schizophrenia, and a minimum of one other was diagnosed with schizophrenia or schizoaffective disorder. These 70 families were the first of a group of 100 that had all final diagnostic information completed and enough DNA available to the laboratory for a complete genome scan. Of the 100, 30 were eliminated due to genotyping problems or failure to contain one sib with schizophrenia only. A larger group of over 300 families was part of the cohort collected from the above sites. However, these were not yet ready for analyses when this scan began approximately 2 years ago. DNA screening of the total group is now in progress. All subjects signed written informed consent for participation in these studies.

Diagnoses were made using DSM-III-R criteria, based on a combination of a structured modified SADS interview (Schedule for Affective Disorder and Schizophrenia [Spitzer and Endicott, 1978]) combined with the SIDP (Structured Interview for Personality Disorders [Pfohl et al., 1990]) or the DIGS (Diagnostic Interview for Genetic Studies [Nurnberger et al., 1994]), a structured questionnaire asked of reliable family informants about other family members, and medical records as indicated. The physicians and other professionals performing clinical evaluations (approximately 2 individuals per site) had been trained in these procedures by one of us (L.E.D.), and all had undergone periodic diagnostic reliability exercises to maintain consistency between centers. Final diagnoses were made by consensus in the UK by J.L., L.E.D., and T.J.C., in Belgium by M.D.H. and L.E.D., in Italy by M.C. and A.V., and in the US by L.E.D., G.S., and M.K. All of the above individuals underwent periodic diagnostic reliability exercises, with kappa statistic scores ranging between 0.85–0.90 for primary diagnoses [Garner et al., 1996].

Pedigrees for each family were diagrammed and all families received research code numbers and separate numbers for each individual. Computer diagnostic files were maintained without knowledge of laboratory marker data. Similarly, laboratory marker data were entered into the file without knowledge of diagnostic information. The families consisted of affected sib-pairs with either schizophrenia or schizoaffective disorder. Thirty-eight families had both parents genotyped, 27 had one parent genotyped, and 5 families had no parents genotyped.

Three phenotype classifications were used in the analysis. The most stringent phenotype contained pedigrees in which affected individuals with schizophrenia only were used. This category contained 48 pedigrees, 115 affected individuals, and 70 ill sib-pairs. The second category contained pedigrees in which affected individuals with schizophrenia or schizoaffective disorder were used. This group resulted in 21 additional pedigrees, for a total of 70 pedigrees containing 171 affected individuals and 101 sib-pairs. The final category contained affected individuals with a broad spectrum of diagnoses including schizophrenia, schizoaffective disorder, paranoid or schizotypal personality disorder, psychosis not otherwise specified (NOS), delusional disorder, and brief reactive psychosis. This group expanded the number of affecteds to 182 individuals (111 ill sib-pairs). Members of some pedigrees also had individuals with affective disorder diagnoses. In order to insure that a more homogeneous phenotype was being tested, pedigrees in both the schizophrenia plus schizoaffective and broad categories were only included if the pedigree had at least one individual affected with schizophrenia. The number of pedigrees broken down by affected sibship size and phenotype is shown in Table II. In addition, the total number of affected sib-pairs for each phenotype is given. For pedigrees with greater than 2 affected sibs, all possible combinations of sibs were counted ($n(n-1)/2$; n = sibship size), and therefore all affected sib-pairs counted are not independent. Twelve of the 70 families were extended beyond the nuclear family to ascertain ill relatives in other branches. Two families in the broad diagnosis category are extended families and each has 2 sibships with ill individuals, accounting for a total of 72 sibships in 70 pedigrees.

Genotyping and Map Construction

Multiplex fluorescent-based genotyping was carried out, using methods previously described [Hall et al., 1996]. Three hundred and thirty-eight highly polymorphic microsatellite markers were chosen primarily from markers developed by Genethon [Dib et al., 1996], the Cooperative Human Linkage Center (CHLC) [1994], or the Utah Marker Development Group [1995]. All markers are described in the Genome Database (GDB). Markers were chosen to be genotyped as part of the genome scan based on location, heterozygosity, PCR quality, and the ability to be multiplexed. Genotype errors as a result of non-Mendelian segregation in pedigrees were detected as described in Hall et al. [1996]. Individual genotypes with obvious errors were discarded. When more ambiguous errors occurred, i.e., it was not clear which family member contained the genotyping error, the whole pedigree was discarded for each particular marker. Two pedigrees consistently contained errors and were therefore discarded for all

TABLE II. Number of Pedigrees Containing Various Affected Sibship Sizes and Total Sample Size Per Phenotype

	Number of affected sibs per pedigree				Total numbers	
Phenotype	2	3	4	5	Pedigrees	ASP[a]
Schizophrenia	42	4	1	1	48	70
Schizophrenia + schizoaffective	61	6	2	1	70	101
Broad	59	10	2	1	72	111

[a]Affected sib-pairs (all combinations, nonindependent).

markers (these are not included in the total number of 70). Overall, 988 errors were detected out of a total of 160,304 genotypes generated in this genome scan, resulting in an error rate of approximately 0.6%.

Genetic maps were constructed for each chromosome using genotype data from the above 70 pedigrees. Family genotype data were used instead of the available CEPH genotype data to construct genetic maps because of the increased number of informative meioses in our data set. On average, there were 350 informative meioses for each marker, although most meioses were phase-unknown since few grandparents were available for genotyping in this family collection (average number of phase-known meioses per marker = 12). Genetic maps for each chromosome were constructed using the automated mapping program MULTIMAP [Matise et al., 1994]. The resulting distances between markers are given in centimorgans using the Kosambi map function, and are similar to those determined using CEPH reference pedigrees.

Statistical Analysis

Power analyses were calculated to determine the likelihood of detecting linkage under various gene effect sizes, using sample sizes corresponding to the three phenotype categories in this study. The methods used to determine the estimated power were the same as those described in Risch [1990]. Since these power calculations assume fully informative matings and the average heterozygosity of our markers was approximately 75%, these values represent maximum power. In addition, since markers in this study were spaced in approximate 10-cM intervals, the power to detect linkage was calculated at a recombination fraction of 0.05 between marker and disease loci (on average, the maximum distance between a potential disease locus to a marker locus). These power calculations are shown in Table III. Using a maximum lod score (MLS) threshold of 2, a reasonable threshold for a first-pass genome scan, our sample sizes exceeded the standard acceptable level of power (0.8) for genes with a larger effect (lambda-s = 4) for both the schizophrenia plus schizoaffective disorder and broad phenotype categories. The schizophrenia-only phenotype still has limited power at this gene-effect size (power = 0.62).

All of the analyses were conducted using each of the three diagnostic categories described (schizophrenia only, schizophrenia plus schizoaffective disorder, and broad diagnosis). Allele frequencies for each marker were calculated from the family genotype data. In regions demonstrating some evidence for linkage (Table IV), allele frequencies were varied and results from the analyses were compared. No significant differences in the results were observed.

For each pedigree, multipoint non-parametric lod (NPL) scores and corresponding P values were computed at each marker and estimated at five equally spaced positions between each marker on the chromosome, using the GENEHUNTER program [Kruglyak et al., 1996]. Map order and distances were used as estimated from the genetic maps constructed for each chromosome. In addition, NPL scores and corresponding P values were estimated 5 cM beyond the first and last markers, respectively, of the multipoint analyses at 1-cM increments. The multipoint NPL scores and corresponding P values were summed over all pedigrees.

Two-point affecteds-only parametric analyses were conducted for all markers on 12 chromosomes (1, 2, 4, 5, 8, 10, 11, 12, 13, 14, 16, and 22), using the program MLINK of FASTLINK [Lathrop et al., 1984; Cottingham et al., 1993]. These chromosomes had at least one marker with a corresponding P value <0.05, using multipoint nonparametric analysis. Both dominant and recessive models were tested for each marker. The parameters for these models were identical to those described by Straub et al. [1995]. Both dominant and recessive models had a reduced penetrance of 0.55, a phenocopy rate of 0.0006, and disease allele frequencies of 0.0049 and 0.0991 for dominant and recessive models, respectively. Lod scores were computed for each pedigree and summed over all pedigrees. The recombination fraction (theta) was varied between 0–0.5 in increments of 0.02 until the lod score was maximized. In addition, summed lod scores for each marker were tested for evidence of heterogeneity, using the program HOMOG [Ott, 1991]. Lod scores computed under heterogeneity (HLOD) were estimated by varying values of alpha (proportion of linked families) and theta until the HLOD scores were maximized. For parametric analyses, MLINK was chosen over GENEHUNTER so that Z_{max} and corresponding theta values could be estimated rather than calculating a lod score at theta = 0.

RESULTS

In order to assess genome coverage and determine accurate genetic distances between markers used in the genome scan, genetic maps were constructed from the genotype data generated from the set of 70 pedigrees. The estimated genome length from this data set

TABLE III. Power of Various Sample Sizes Corresponding to Sz (Schizophrenia), Sz + SA (Schizophrenia + Schizoaffective), and Broad Phenotype Categories, Respectively, Assuming a Recombination Fraction of 0.05 Between Marker and Disease Loci, and Full Information Content*

MLS threshold	λ-s = 2, no. of ASP			λ-s = 3, no. of ASP			λ-s = 4, no. of ASP		
	70	101	111	70	101	111	70	101	111
2	0.16	0.25	0.30	0.44	0.62	0.71	0.62	0.81	0.87
3	0.04	0.10	0.14	0.17	0.38	0.48	0.30	0.59	0.71

*ASP, affected sib-pairs.

TABLE IV. Markers With Parametric lod Scores Above 1.5, or Peak Nonparametric Multipoint NPL Scores With Corresponding *P* Values < 0.05 and Their Corresponding Phenotype

Marker	Chromosome location	Distance from pter (cM)	Phenotype	Z_{max}	Theta	HLOD	Theta	Alpha	NPL (*P* <)
Chromosome 1									
D1S551	1p31–p13	118.4	Broad	1.41	0.20[a]	1.41	0.20	1.00	1.65 (.05)
D1S196	1q22–q23	184.1	Sz + SA	1.81	0.18[b]	2.40	0.00	0.46	1.04 (.14)
D1S1599	1q31–q32	207.0	Broad	0.46	0.28[b]	0.57	0.04	0.22	1.86 (0.03)
Chromosome 2									
D2S1337	2p14–p13	60.2	Sz + SA	1.56	0.16[a]	2.19	0.00	0.41	1.97 (.02)
			Broad	1.60	0.18[a]	1.98	0.00	0.36	2.13 (.01)
Chromosome 4									
D4S1546	4pter–cen	18.9	Sz	1.91	0.12[a]	2.44	0.00	0.56	1.35 (.05)
			Sz + SA	1.22	0.20[a]	2.05	0.00	0.40	1.96 (.02)
D4S2456	4q13–q21	65.3	Sz	1.89	0.12[a]	1.89	0.12	1.00	1.59 (.03)
D4S400	4q11–q21	68.2	Sz + SA	1.48	0.16[a]	1.94	0.00	0.50	1.75 (.03)
Chromosome 5									
D5S406	5p15	0.0	Sz + SA	0.85	0.26[b]	0.85	0.26	1.00	2.18 (.01)
D5S433	5p14–q21	109.8	Sz + SA	0.91	0.26[b]	1.00	0.08	0.33	1.60 (.05)
Chromosome 8									
D8S560	8p22–p21	36.0	Sz	1.99	0.16[b]	1.99	0.16	1.00	1.20 (.07)
GAAT1A4	8q22	103.7	Broad	0.37	0.30[b]	0.37	0.30	1.00	1.58 (.05)
Chromosome 10									
D10S677	10q22–q24	113.8	Sz + SA	0.74	0.26[b]	0.74	0.20	0.60	2.08 (.01)
Chromosome 11									
D11S1393	11pter–p11	59.0	Sz	2.30	0.10[a]	2.40	0.00	0.59	0.89 (.14)
D11S2002	11q21–q23	78.2	Sz	0.88	0.26[b]	1.59	0.00	0.28	1.59 (.03)
Chromosome 12									
D12S85	12p13–q24	45.9	Sz + SA	0.34	0.26[a]	0.34	0.20	0.64	2.01 (.02)
D12S342	12q22–q24	131.4	Broad	0.82	0.24[a]	0.82	0.00	0.28	1.57 (.05)
Chromosome 13									
D13S1293	13q12–q13	16.5	Sz	2.85	0.10[a]	2.85	0.10	1.00	1.80 (.01)
D13S168	13q14	35.6	Sz	1.56	0.16[a]	1.72	0.00	0.49	1.51 (.03)
D13S170	13q31	65.8	Sz + SA	0.63	0.28[a]	1.25	0.00	0.25	1.83 (.03)
D13S1315	13q34	108.4	Broad	0.58	0.28[a]	0.64	0.00	0.23	1.69 (.04)
Chromosome 14									
D14S290	14q21–q23	51.0	Sz + SA	0.02	0.38[b]	0.45	0.00	0.17	1.99 (.02)
D14S1012	14q21–q23	54.4	Sz + SA	1.35	0.24[b]	1.99	0.00	0.29	1.71 (.04)
Chromosome 16									
D16S421	16q22.1	62.7	Sz	0.12	0.32[b]	0.21	0.00	0.17	2.09 (.01)
Chromosome 22									
D22S446	22q11	13.8	Sz + SA	1.50	0.22[b]	1.50	0.22	1.00	2.16 (.01)
D22S283	22q12–q13	42.0	Sz	1.77	0.20[b]	1.77	0.20	1.00	1.64 (.02)

[a]Dominant.
[b]Recessive.

was 3,310.3 cM, providing an average resolution of 10.5 cM between each of the 338 markers, with an interval range of 0–31 cM. Most intervals between markers were between 5–20 cM (228 out of 315 intervals; 72%). Seven intervals had no recombination between markers. Markers chosen for the genome scan were highly polymorphic microsatellite markers, with an average observed heterozygosity of 74.3%.

Each of the 338 markers was analyzed using multipoint nonparametric allele-sharing tests in GENEHUNTER for each of the three phenotype classifications described, using map order and distances estimated from the family genotype data. Figure 1 shows the multipoint curves for each chromosome. Chromosomes 1, 2, 4, 5, 8, 10, 11, 12, 13, 14, 16, and 22 have at least one peak NPL score with a corresponding *P* value <0.05 (see Table IV). These *P* values are uncorrected for genome-wide tests. The most significant peak from the nonparametric multipoint analyses occurred at D16S421 located in 16q22.1, using a schizophrenia-only phenotype (NPL = 2.09, $P < 0.01$).

The information content was also calculated using GENEHUNTER at the same positions as the multipoint analyses, i.e., for all markers along the chromosome and in five equally spaced positions between each marker. The average information content for all autosomal markers was 0.67 (range, 0.52–0.78). The information content curves for each chromosome are shown in Figure 2.

In order to obtain more evidence for the suggestive regions found using nonparametric linkage analysis, parametric tests were conducted on the 12 chromosomes demonstrating evidence for linkage. Both dominant and recessive affecteds-only models were used under each of the three diagnostic phenotypes described, using MLINK of FASTLINK. In addition to calculating the lod score, an HLOD and corresponding alpha value was also calculated using HOMOG in order to determine the lod score under heterogeneity and the proportion of families linked to a particular marker, respectively. Nine of the 12 chromosomes analyzed had at

TABLE V. All Loci Examined and Corresponding Highest Multipoint NPL Score Obtained

Marker	Distance from pter (cM)	Phenotype	NPL	P<
Chromosome 1				
D1S243	0.0	SZ	−0.08	0.54
D1S548	9.2	Broad	0.62	0.25
D1S228	23.6	Broad	0.64	0.25
D1S552	38.7	Sz + SA	0.51	0.29
D1S164	56.7	Sz + SA	1.51	0.06
D1S3175	77.7	Broad	1.44	0.07
D1S232	82.7	Broad	0.80	0.20
D1S405	90.2	Broad	0.45	0.32
D1S209	103.0	Broad	−0.06	0.52
D1S192	107.0	Broad	0.49	0.30
D1S551	118.4	Broad	1.65	0.04
D1S167	128.1	Broad	1.34	0.08
D1S223	139.7	Broad	0.77	0.21
D1S453	159.1	Sz	0.12	0.44
D1S484	171.9	Sz + SA	0.49	0.30
D1S194	178.1	Sz + SA	0.55	0.28
D1S196	184.1	Sz + SA	1.04	0.14
D1S466	202.6	Broad	1.06	0.13
D1S240	205.4	Broad	1.79	0.03
D1S1599	207.0	Broad	1.86	0.03
D1S245	229.3	Sz	1.36	0.05
D1S1602	234.5	Sz	1.35	0.05
D1S235	256.7	Sz	0.72	0.18
D1S517	266.3	Broad	0.44	0.32
D1S2842	276.9	Sz	0.41	0.30
Chromosome 2				
D2S1329	0.0	Sz	0.63	0.22
D2S165	25.2	Sz	1.13	0.08
D2S367	35.5	Broad	0.88	0.18
D2S391	50.0	Broad	1.43	0.07
D2S1337	60.2	Broad	2.13	0.01
D2S438	76.4	Broad	1.55	0.05
D2S139	82.5	Sz + SA	1.57	0.05
D2S135	91.6	Sz + SA	1.44	0.06
D2S2224	112.1	Sz + SA	1.40	0.07
D2S347	113.1	Sz	1.07	0.09
D2S114	124.9	Broad	1.11	0.12
D2S368	126.1	Broad	1.09	0.13
D2S150	130.8	Broad	0.55	0.28
D2S141	144.3	Sz	0.14	0.43
D2S93	145.4	Sz	0.51	0.26
D2S418	145.5	Sz	0.52	0.26
D2S326	157.6	Sz	0.27	0.37
D2S389	169.2	Sz	−0.39	0.69
D2S115	175.4	Sz + SA	0.06	0.47
D2S116	176.8	Sz + SA	−0.06	0.52
D2S422	181.5	Sz + SA	−0.18	0.57
D2S434	192.2	Sz	0.01	0.49
D2S126	196.3	Sz	−0.13	0.56
D2S439	206.7	Sz	0.50	0.26
D2S206	212.0	Sz	0.55	0.24
D2S338	230.1	Broad	1.02	0.14
Chromosome 3				
D3S1307	0.0	Sz	0.13	0.43
D3S1297	2.5	Sz	0.06	0.46
D3S1304	13.9	Sz	0.42	0.30
D3S1597	18.4	Sz	0.22	0.39
D3S1286	28.9	Sz + SA	0.23	0.40
D3S2396	45.6	Broad	−0.53	0.71
D3S1768	52.3	Sz	−0.10	0.55
D3S1611	52.4	Sz	−0.11	0.55
D3S1300	73.4	Broad	0.21	0.41
D3S3571	82.3	Broad	0.11	0.45
D3S1598	93.7	Sz	−0.27	0.63
D3S2406	95.6	Sz	−0.76	0.83
D3S2388	99.5	Sz	−0.83	0.85
D3S1276	100.3	Sz	−0.74	0.82
D3S2386	100.1	Sz	−0.78	0.83
D3S1271	103.7	Sz	−0.61	0.77
D3S1610	113.8	Sz	−0.23	0.61
D3S1303	119.7	Sz	−0.48	0.72
D3S1292	131.3	Broad	0.76	0.21
D3S1569	143.3	Broad	0.87	0.18
D3S196	146.3	Broad	1.32	0.09
D3S1282	161.6	Broad	−0.02	0.51
D3S1754	177.9	Sz + SA	0.27	0.38
D3S2398	199.1	Sz + SA	1.28	0.09
D3S1314	201.1	Sz + SA	1.07	0.13
Chromosome 4				
D4S394	0.0	Sz + SA	0.54	0.28
D4S1546	18.9	Sz + SA	1.96	0.02
D3A632	32.5	Sz + SA	0.31	0.37
D4S2382	42.9	Sz	−0.33	0.66
D4S2379	54.3	Sz	0.16	0.42
D4S2456	65.3	Sz + SA	1.63	0.04
D4S400	68.2	Sz + SA	1.75	0.03
D4S1647	89.7	Broad	1.40	0.07
D4S1651	103.1	Sz	0.68	0.20
D4S420	121.0	Broad	−0.12	0.55
D4S1625	124.7	Sz + SA	−0.02	0.50
D4S413	136.5	Broad	0.32	0.36
D4S2368	143.7	Broad	0.40	0.33
D4S415	154.8	Broad	1.20	0.10
D4S408	168.1	Broad	1.57	0.05
Chromosome 5				
D5S406	0.0	Sz + SA	2.18	0.01
D5S807	9.9	Broad	1.49	0.06
D5S416	19.9	Broad	0.68	0.23
D5S268	23.9	Broad	0.46	0.31
D5S477	37.4	Broad	−0.23	0.59
D5S407	56.9	Sz	−0.04	0.52
D5S424	78.3	Sz + SA	0.39	0.33
D5S401	95.0	Sz + SA	1.09	0.12
D5S433	109.8	Sz + SA	1.60	0.05
D5S804	126.8	Sz	−0.19	0.59
D5S399	135.8	Sz + SA	0.19	0.42
FGFA	141.3	Sz + SA	0.05	0.48
D5S2112	156.8	Sz + SA	−0.75	0.79
D5S1475	170.1	Sz	0.24	0.38
D5S429	183.1	Sz	0.61	0.22
D5S498	191.0	Sz	0.80	0.16
Chromosome 6				
D6S344	0.0	Sz	−0.16	0.57
D6S277	15.1	Sz	−0.63	0.78
D6S1006	21.2	Sz	−0.96	0.88
D6S399	21.3	Sz	−0.95	0.88
D6S422	35.9	Broad	0.43	0.32
D6S291	50.1	Sz	0.19	0.40
D6S426	57.4	Sz	0.43	0.29
D6S1018	64.1	Sz	0.59	0.23
D6S493	80.7	Sz	0.28	0.36
D6S501	96.0	Sz	−0.22	0.61
D6S474	111.3	Broad	0.48	0.30
D6S1009	130.7	Sz + SA	1.34	0.08
D6S495	146.0	Sz + SA	0.67	0.24
D6S1008	162.1	Sz + SA	0.82	0.19
D6S281	180.4	Sz	−0.69	0.80
Chromosome 7				
D7S513	0.0	Sz	−0.06	0.52
D7S493	13.2	Sz	−0.51	0.73
D7S484	30.0	Sz	0.10	0.44
D7S645	52.1	Sz + SA	−0.29	0.62
D7S1511	52.9	Broad	−0.07	0.52
D7S1797	61.6	Broad	−0.64	0.75
D7S524	66.5	Sz	−1.27	0.95
D7S1789	71.2	Broad	−0.57	0.72

TABLE V. All Loci Examined and Corresponding Highest Multipoint NPL Score Obtained

Marker	Distance from pter (cM)	Phenotype	NPL	P<
D7S518	80.5	Broad	0.85	0.18
D7S677	92.7	Broad	1.49	0.06
D7S1809	95.1	Broad	1.36	0.08
D7S530	98.9	Broad	1.18	0.11
D7S640	106.3	Broad	0.87	0.18
D7S794	118.7	Broad	1.10	0.12
D7S1807	139.1	Sz	0.57	0.24
Chromosome 8				
D8S264	0.0	Broad	0.47	0.31
D8S265	17.5	Sz	−0.39	0.68
D8S552	21.2	Sz	−0.42	0.69
D8S560	35.9	Sz	1.20	0.07
D8S282	38.8	Sz	1.19	0.07
D8S1104	58.6	Sz	0.77	0.17
D8S1102	61.7	Sz	0.78	0.17
D8S260	67.6	Sz	1.33	0.05
D8S286	82.5	Sz	0.68	0.20
GAAT1A4	103.5	Broad	1.58	0.05
D8S281	113.8	Broad	1.03	0.14
D8S568	125.8	Sz	0.93	0.13
D8S272	142.4	Sz	0.53	0.25
D8S373	162.0	Sz	0.31	0.35
Chromosome 9				
D9S286	0.0	Broad	1.16	0.11
D9S156	15.9	Broad	0.77	0.21
D9S741	28.1	Broad	0.34	0.35
D9S169	29.3	Broad	0.43	0.32
D9S50	40.9	Broad	0.93	0.16
D9S301	46.5	Broad	1.19	0.10
D9S153	56.7	Sz + SA	0.55	0.28
D9S257	68.2	Broad	1.06	0.13
D9S278	71.4	Broad	1.02	0.14
D9S176	79.6	Sz + SA	0.78	0.20
D9S299	86.3	Sz + SA	1.02	0.14
D9S170	106.5	Sz	0.26	0.37
D9S159	118.7	Sz	0.83	0.15
D9S66	134.5	Sz	0.61	0.22
Chromosome 10				
D10S1212	0.0	Sz	0.63	0.21
D10S249	0.1	Sz	0.66	0.20
D10S179	15.8	Sz	0.80	0.16
D10S547	24.9	Broad	1.23	0.10
D10S1214	29.1	Broad	1.27	0.09
D10S191	35.8	Broad	1.37	0.08
D10S611	52.1	Sz	0.74	0.18
D10S601	55.6	Sz	0.89	0.13
D10S1426	56.0	Sz	0.82	0.15
D10S220	66.0	Sz	1.27	0.06
D10S1211	77.7	Sz + SA	1.39	0.07
D10S537	83.9	Sz + SA	1.30	0.08
D10S535	87.6	Sz + SA	1.10	0.12
D10S2327	94.7	Sz + SA	1.62	0.04
D10S677	113.8	Sz + SA	2.08	0.01
D10S597	122.1	Broad	1.45	0.06
D10S670	122.2	Broad	1.44	0.07
D10S610	134.1	Broad	0.66	0.24
D10S587	148.4	Broad	−0.23	0.59
D10S1655	165.6	Sz	−0.31	0.65
Chromosome 11				
D11S1984	0.0	Sz	0.82	0.16
D11S1999	16.5	Sz	0.74	0.18
D11S419	25.3	Sz	0.53	0.26
D11S902	28.2	Sz	0.38	0.32
D11S1888	29.1	Sz	0.35	0.33
D11S1322	47.1	Sz + SA	1.04	0.14
D11S1301	48.0	Sz + SA	1.07	0.13
D11S1968	48.1	Sz + SA	1.06	0.13
D11S4200	49.5	Sz + SA	0.62	0.25

TABLE V. (Continued)

Marker	Distance from pter (cM)	Phenotype	NPL	P<
D11S1393	59.0	Sz	0.89	0.14
D11S1385	61.8	Sz	0.54	0.25
D11S1335	62.2	Sz	0.54	0.25
D11S1298	62.3	Sz	0.54	0.25
D11S2002	78.2	Sz	1.59	0.03
D11S1367	84.5	Broad	1.35	0.08
D11S35	92.3	Broad	1.21	0.10
D11S1391	101.7	Broad	0.88	0.18
D11S924	114.7	Broad	−0.01	0.50
D11S925	117.6	Sz + SA	0.11	0.45
D11S4150	135.1	Sz	1.09	0.09
D11S968	149.2	Sz	0.93	0.13
Chromosome 12				
D12S1725	0.0	Broad	0.44	0.32
D12S77	10.7	Sz	0.86	0.14
D12S364	18.8	Broad	1.82	0.03
D12S373	21.1	Broad	1.87	0.03
D12S87	37.3	Broad	1.95	0.02
D12S85	45.9	Sz + SA	2.01	0.02
D12S297	53.1	Sz + SA	0.93	0.16
D12S398	54.1	Sz + SA	0.71	0.22
D12S83	59.5	Sz + SA	1.05	0.13
D12S375	69.7	Sz + SA	1.18	0.11
D12S1717	86.1	Broad	0.38	0.34
PAH	98.3	Sz	0.40	0.30
D12S1074	99.7	Sz	0.30	0.35
D12S1091	104.0	Sz	−0.06	0.53
D12S79	116.4	Broad	1.08	0.13
D12S378	127.9	Broad	1.39	0.07
D12S342	131.4	Broad	1.57	0.05
D12S343	151.4	Broad	1.08	0.13
Chromosome 13				
D13S283	0.0	Sz + SA	0.77	0.21
D131293	16.5	Sz	1.80	0.01
D13S220	18.2	Sz	1.72	0.02
D13S325	27.3	Sz	1.56	0.03
D13S328	33.6	Sz	1.25	0.06
D13S168	35.6	Sz	1.51	0.03
D13S318	53.5	Sz + SA	0.49	0.30
D13S170	65.8	Sz + SA	1.83	0.03
D13S317	67.4	Sz + SA	1.63	0.04
D13S762	74.0	Sz + SA	1.34	0.08
D13S158	91.8	Broad	0.60	0.26
D13S173	101.6	Broad	1.10	0.12
D13S1315	108.4	Broad	1.69	0.04
D13S285	116.4	Broad	1.59	0.05
D13S293	127.2	Broad	0.78	0.20
Chromosome 14				
D14S50	0.0	Sz + SA	1.64	0.04
D14S264	10.1	Sz + SA	0.76	0.21
D14S70	24.8	Sz + SA	0.20	0.41
D14S278	32.7	Sz + SA	0.78	0.20
D14S269	39.3	Sz + SA	0.95	0.15
D14S290	51.0	Sz + SA	1.99	0.02
D14S1012	54.4	Sz + SA	1.71	0.04
D14S74	72.1	Sz + SA	1.31	0.08
D14S606	75.9	Sz + SA	1.07	0.13
D14S280	86.9	Sz	0.53	0.25
D14S302	94.1	Sz	0.55	0.24
D14S267	108.2	Sz	1.44	0.04
D14S1010	120.5	Sz	0.94	0.12
Chromosome 15				
D15S128	0.0	Sz	0.25	0.37
D15S144	17.5	Sz	0.89	0.13
D15S641	26.3	Sz	0.92	0.12
D15S659	36.5	Sz	0.78	0.16
D15S148	46.2	Sz	1.02	0.10
D15S159	56.4	Sz	0.57	0.24
D15S131	64.2	Sz	0.55	0.24

TABLE V. All Loci Examined and Corresponding Highest Multipoint NPL Score Obtained

Marker	Distance from pter (cM)	Phenotype	NPL	P<
D15S152	71.9	Sz	0.17	0.41
D15S130	93.3	Broad	0.25	0.39
D15S207	101.2	Broad	0.16	0.43
D15S642	121.9	Broad	0.99	0.15
Chromosome 16				
D16S407	0.0	Broad	−0.61	0.74
D16S748	6.1	Sz + SA	−0.64	0.75
D16S500	10.2	Sz	−0.98	0.89
D16S417	23.5	Sz + SA	−0.09	0.54
D16S401	26.0	Sz	0.02	0.49
D16S753	36.8	Sz + SA	0.61	0.26
D16S408	48.3	Sz	1.17	0.07
D16S320	52.0	Sz	1.27	0.06
D16S421	62.7	Sz	2.09	0.01
D16S752	65.0	Sz	1.83	0.01
D16S750	83.9	Sz + SA	0.25	0.39
D16S3037	100.7	Sz	0.52	0.26
Chromosome 17				
D17S1828	0.0	Broad	−0.17	0.57
D17S1303	27.1	Sz	0.01	0.49
D17S969	29.2	Sz	0.39	0.31
D17S799	32.4	Sz	0.74	0.18
D17S250	55.9	Sz	0.17	0.41
D17S808	80.0	Sz	0.06	0.47
D17S948	82.6	Sz	0.20	0.40
D17S1291	89.1	Sz	0.30	0.35
D17S968	106.4	Sz	−0.15	0.57
D17S784	122.3	Broad	0.55	0.28
Chromosome 18				
D18S452	0.0	Sz	0.71	0.19
D18S53	16.1	Sz	1.00	0.11
D18D847	27.2	Sz	0.25	0.38
D18S548	36.8	Sz	0.31	0.35
D18S487	47.7	Sz	0.23	0.38
D18S849	54.7	Sz	0.47	0.28
D18S68	64.9	Sz	0.32	0.34
D18S541	77.0	Sz	−0.28	0.63
D18S1161	83.0	Sz	−0.19	0.59
Chromosome 19				
D19S878	0.0	Sz + SA	0.35	0.35
D19S177	12.7	Sz + SA	0.44	0.32
D19S406	20.5	Sz + SA	1.26	0.09
D19S391	25.7	Sz + SA	0.80	0.20
D19S432	38.2	Broad	−0.29	0.62
D19S434	48.6	Sz	−1.00	0.89
D19S433	51.2	Sz	−0.62	0.78
D19S225	56.4	Sz	−0.64	0.79
D19S412	71.9	Sz	0.42	0.30
D19S601	88.5	Sz	0.80	0.16
D19S418	100.3	Sz	0.77	0.17
D19S254	108.8	Sz	0.53	0.25
Chromosome 20				
D20S889	0.0	Broad	−0.22	0.59
D20S189	24.7	Sz	0.66	0.20
D20S471	41.1	Sz	0.32	0.34
D20S200	52.5	Sz	0.35	0.33
D20S119	67.2	Sz + SA	0.67	0.24
D20S476	79.8	Sz	1.03	0.10
D20S171	101.7	Broad	0.18	0.42
Chromosome 21				
D21S1431	0.0	Sz + SA	0.04	0.48
D21S265	13.7	Sz + SA	0.36	0.35
D21S1435	16.6	Broad	0.48	0.30
D21S167	33.5	Sz + SA	−0.28	0.61
D21S1261	49.5	Broad	−0.47	0.69
Chromosome 22				
F8VWFP	0.0	Sz + SA	1.78	0.03
D22S446	13.8	Sz + SA	2.16	0.01
D22S275	31.7	Sz	1.22	0.06
D22S278	39.6	Sz	1.35	0.05
D22S283	42.0	Sz	1.64	0.02
D22S445	46.8	Sz	0.93	0.12
D22S274	57.1	Sz + SA	0.88	0.17
X chromosome				
DXS8051	0.0	Sz	0.12	0.44
DXS207	17.2	Sz + SA	0.09	0.46
DMD	44.2	Sz + SA	0.34	0.36
MAOB	57.9	Sz + SA	0.24	0.40
DXS1055	66.1	Broad	0.15	0.43
DXS991	74.5	Broad	0.58	0.27
DXS6786	90.5	Broad	0.43	0.32
DXS990	92.9	Sz + SA	0.48	0.30
DXS1191	101.6	Sz + SA	0.54	0.28
DXS1220	107.6	Sz + SA	0.78	0.20
DXS1047	139.1	Sz + SA	0.93	0.16
DXS1227	149.8	Sz + SA	0.59	0.26
DXS1123	168.2	Broad	1.34	0.08

least one marker with an HLOD score greater than 1.5 (Table IV).

DISCUSSION

The results from this initial systematic genome-wide screen of families with schizophrenia do not provide statistically significant evidence for a major genetic locus, but weak positive lod scores were present within as many as 12 chromosomal regions (1q22–q32, 2p14–p12, 4q21–q23, 5q, 8, 10, 11, 12, 13q12–q14, 14q23, 16q13–22, and 22q11).

A possible linkage region on chromosome 1q22–q32 is supported by evidence from markers D1S1599 and D1S196, located approximately 23 cM apart (HLOD = 2.4; alpha = 0.48). No other investigative groups are known to have positive findings in this region.

On chromosome 2p14–12, marker D2S1337 shows a peak multipoint NPL of 2.13 ($P = 0.01$) and maximum HLOD of 2.2 (alpha = 0.41). Moises et al. [1995a] reported linkage to D2S135 ($P = 1.1 \times 10^{-6}$) located in 2q12–q14, approximately 30 cM away from the region detected in this scan, and Byerley et al. (unpublished data) have lod scores suggestive of linkage within this region in a cohort of families with schizophrenia from the island of Palau.

Another region of potential interest is located on chromosome 4; positive markers span an approximate 49-cM region: D4S1546 (18.9 cM from pter), D4S2456 (65.4 cM from pter), and D4S400 (68.3 cM from pter). The nonparametric multipoint curve for chromosome 4 indicates two distinct peak regions. The genome scans of Moises et al. [1995a] and Williams et al. [1997] resulted in positive scores in this region.

Straub et al. [1997] recently reported evidence for a schizophrenia susceptibility locus within 5q22–q31. One marker from our present study, D5S433, within this region had an NPL of 1.60, $P < 0.05$. Wildenauer et al. [1997] also provided positive data in this region.

Two groups have reported evidence for a locus within

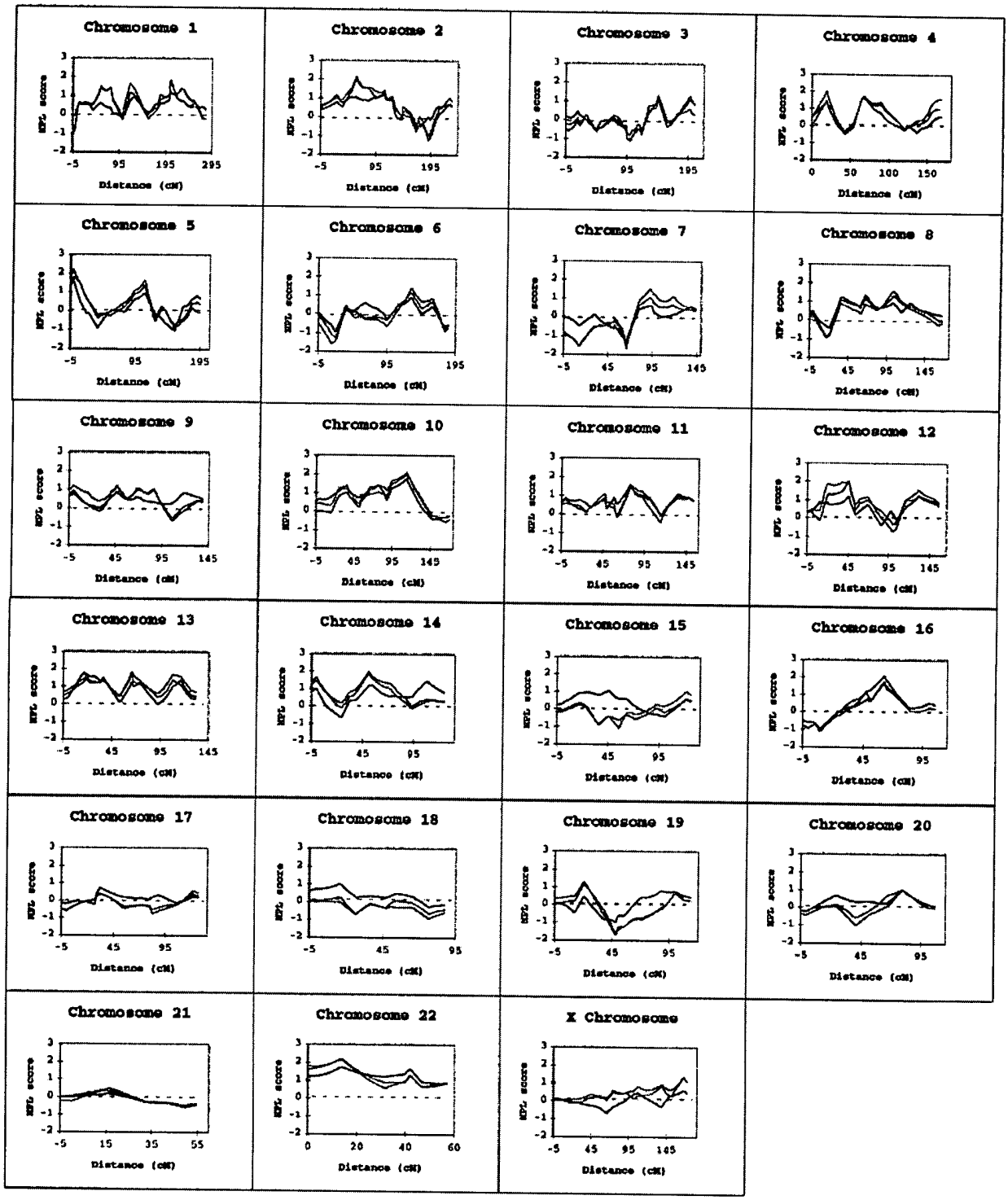


Fig. 1. Multipoint nonparametric linkage analysis of the genome. Three hundred and thirty-eight markers spanning the entire genome were analyzed using multipoint nonparametric allele-sharing tests in GENEHUNTER. Cumulative map distances in centimorgans starting from the first pter marker are given on each X-axis. Multipoint NPL scores totaled over all pedigrees are given on each Y-axis. The resulting multipoint curve using the schizophrenia-only phenotype is shown in blue, the schizophrenia plus schizoaffective disorder phenotype is shown in red, and the broad diagnosis phenotype is shown in green.

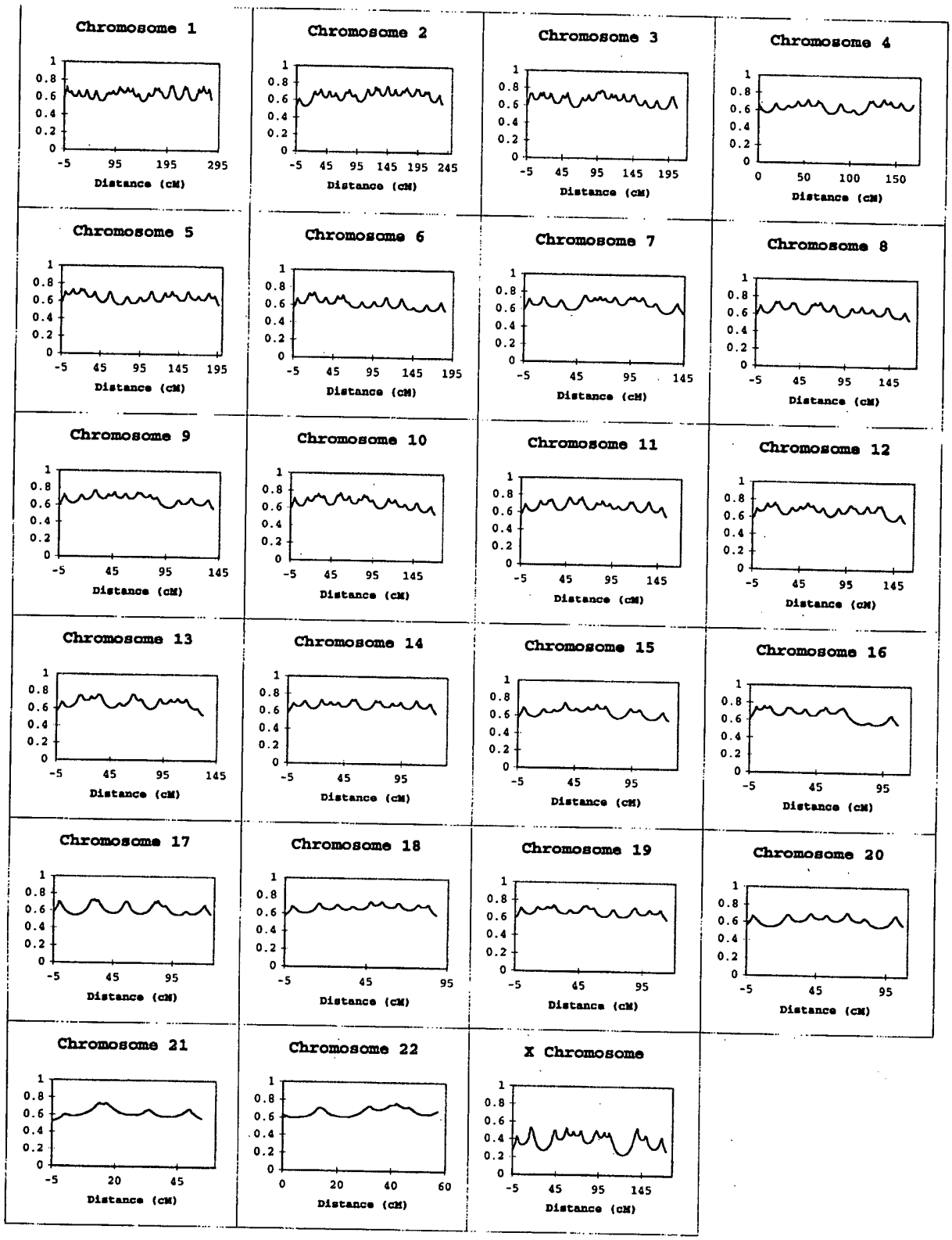


Fig. 2. Information content curve for each chromosome. The information content (Y-axis) at each marker and in five equally spaced intervals between each marker (X-axis) was calculated using GENEHUNTER. The average information content for all autosomal markers was 0.67. The lowest information content (0.52) occurred at the most proximal position on chromosome 21, and the highest information content (0.78) occurred in proximal 11q (distance from pter, 62.3 cM).

8p22–21 [Kendler et al., 1996; Pulver et al., 1995]. One marker in this genome scan was positive in this region (D8S560, Z_{max} = 1.99, theta of 0.16 with recessive inheritance), and similarly, the NIMH Collaborative Initiative for schizophrenia [1997] also noted that this region was of possible interest.

The chromosome 13q12–14 region also provided positive results, specifically with markers D13S1293 (distance from pter, 16.5 cM; schizophrenia-only diagnosis) and D13S168 (distance from pter, 35.6 cM; broad diagnosis). Both nonparametric and parametric scores produced the highest peaks with a dominant model (Z_{max} = 2.85, theta = 0.10). In addition, Lin et al. [1995] reported suggestive evidence for linkage to a locus in 13q14.1–q32, while Antonarakis et al. [1996] reported a positive finding at a region in 13q32. One marker in our study, D13S170, is approximately 15 cM from the reported region in 13q32 and shows slight evidence for linkage using nonparametric analysis (NPL = 1.83, $P < 0.03$).

A possible locus on chromosome 16q13–22 is supported by results from D16S320 and D16S421, located within an approximate 12-cM region. All markers in this region of chromosome 16 support a locus using the schizophrenia phenotype and a recessive mode of inheritance when using parametric analysis. A previous study which contained two overlapping families with the present study [Polymeropoulos et al., 1995] also showed evidence of linkage within this region (sib-pair analysis chi-square = 9.76, $P < 0.0009$) at D16S266.

One of the more interesting positive findings in this search is on chromosome 22 with D22S446 (distance from pter, 13.8 cM) and D22S283 (distance from pter, 42.1 cM). It appears from the nonparametric multipoint analysis that these two linkage regions may be distinct. Carlson et al. [1997] defined a 3-Mb region in 22q11 that is commonly deleted in some psychiatric patients with velo-cardio-facial syndrome (VCFS). D22S426 is within 1 cM of the region defined by Carlson et al. [1997]. Several investigators found evidence for linkage to another region spanning approximately 20 cM in 22q12–q13 [Pulver et al., 1994a; Coon et al., 1994a; Gill et al., 1996; Moises et al., 1995b], which is the same region detected by Marker D22S283. Some groups have subsequently failed to replicate these findings in separate studies [Kalsi et al., 1995; Pulver et al., 1994b; Williams et al., 1997; Parsian et al., 1997]. Our own previous study of this region [Polymeropoulos et al., 1995] was essentially negative. Ten of the 104 families from the study of Polymeropoulos et al. [1995] overlapped with those of the present study. Analysis without these families did not make a difference in any of the reported chromosome 22 results.

With regard to the X chromosome, this genome-wide scan fails to support our previously reported weak positive findings on proximal Xp [DeLisi et al., 1994; Dann et al., 1997]. However, the power available in 70 families may not be sufficient to detect linkage to X-Y homologous genes.

In summary, there are several regions that need closer examination based on the results of the present genome screen for a schizophrenia susceptibility locus (or loci). These are within chromosomes 1, 2, 4, 5, 8, 13, 16, and 22, and some have received further support from other independent genome-wide screens. Nevertheless, none of these regions reach conventional levels for statistical evidence of linkage, as proposed by Lander and Kruglyak [1995], whose guidelines for "significant linkage" in genome-wide scans indicate that a P value of less than 2×10^{-5} or a lod score greater than 3.6 is needed. Thus, it must be considered that all findings in this genome scan could be the result of chance, and that none of the loci discussed actually represent true schizophrenia susceptibility loci. This possibility is increased by the fact that we have analyzed multiple phenotype classifications, which further influence the type I error rate.

One interpretation of the results from this study and previous studies reported to date is that one major locus is unlikely to exist for schizophrenia, since many families have already been screened with no lod score reaching conventional levels of significance. However, if a gene of minor effect is to be found, the power from 70 predominantly nuclear families is too low here for significant results to be likely to appear (see Table III). Sawcer et al. [1997] suggested that the guidelines of Lander and Kruglyak [1995] are too stringent for genome scans searching for complex disease genes, such as schizophrenia.

Regardless of one's definition of "linkage" and whether there is one or several genes involved, this genome-wide search has provided a few regions worth further investigation. The present study is now being extended in the laboratory to densely cover these regions with more markers in approximately 1–3-cM intervals and to expand the number of families at least 3-fold who will receive systematic genomic screening. Replication of similar findings by other independent investigative groups will also aid in validating the positive scores from the present study.

Finally, the controversy over phenotypic characteristics corresponding to the genotype has not been clarified by the present study. No diagnostic inclusion criteria consistently yielded maximal scores for each suggestive locus; rather, the peak lod scores varied among markers for different diagnostic categories and types of analyses. While it is possible that different loci may contribute to different diagnostic subtypes, it is doubtful that this dilemma can be solved until a definite linkage to a locus with an associated mutant variant has been determined. Families for the present initial study were chosen conservatively so that at least one member of a sib-pair had a nonaffective form of schizophrenia (without any depressive or manic components). However, when all schizoaffective individuals were eliminated from analyses, the power for detecting linkage significantly lowered; thus it cannot be determined whether the reduced power or the lack of accurately estimating the phenotypic boundaries is contributing to the variable results.

REFERENCES

Antonarakis SE, Blouin J-L, Pulver AE, Wolyniec P, Lasseter VK, Nestadt GK, Kasch L, Babb R, Kazazian HH, Dombroski B, Kimberland M, Ott J, Housman D, Karayiorgou M, MacLean CJ (1995): Schizophrenia susceptibility and chromosome 6p24–22. Letter to the Editor. Nat Genet 11:235–236.

Antonarakis SE, Blouin JL, Curran M, Luebbert H, Kazazian HH, Dombrioski B, Housman D, Ton C, Karayiorgou M, Chakravarti A, Wolyniec P, Lasseter VK, Nestadt G, Pulver AE (1996): Linkage and sib-pair analysis reveal a potential schizophrenia susceptibility gene on chromosome 13q32. Psychiatr Genet 6:136.

Barr CL, Kennedy JL, Pakstis L, Wetterberg L, Sjogren B, Bierut L, Wadelius C, Wahlstrom J, Martinsson T, Giuffra L, Gelernter J, Hallmeyer J, Moises HW, Kurth J, Cavalli-Sforza LL, Kidd KK (1994): Progress in a genome scan for linkage in schizophrenia in a large Swedish kindred. Am J Med Genet 54:51–58.

Byerley W, Polloi A, Dale P, Hoff M, Rosenthal J, Tiobech J, Reimherr F, Paul W, Myles-Worsley M, Coon H (1997): Genome-wide linkage analysis of a large schizophrenia pedigree ascertained from Palau, Micronesia. Am J Med Genet 74:559–560.

Carlson C, Papolos D, Pandita RK, Faedda G, Veit S, Goldberg R, Shprintzen R, Kucherlapati R, Morrow B (1997): Molecular analysis of velocardio-facial syndrome patients with psychiatric disorders. Am J Hum Genet 60:851–859.

Coon H, Holik J, Hoff M, Reimberr F, Wender P, Freedman R, Byerley W (1994a): Analysis of chromosome 22 markers in 9 schizophrenia pedigrees. Am J Med Genet 54:72–79.

Coon H, Jensen S, Holik J, Hoff M, Myles-Worsley M, Reimherr F, Wender P, Waldo M, Freedman R, Leppert M, Byerley W (1994b): Genomic scan for genes predisposing to schizophrenia. Am J Med Genet 54:59–71.

Cooperative Human Linkage Center (1994): A comprehensive human linkage map with centimorgan density. Science 265:2049–2054.

Cottingham R, Idury R, Schaffer A (1993): Faster sequential genetic linkage computations. Am J Hum Genet 53:252–263.

Crowe RR, Black DW, Wesner R, Andreasen NC, Cookman A, Roby J (1991): Lack of linkage to chromosome 5q11–q13 markers in six schizophrenia pedigrees. Arch Gen Psychiatry 48:351–361.

Dann J, DeLisi LE, Devoto M, Laval S, Nancarrow DJ, Shields G, Smith A, Loftus J, Peterson P, Vita A, Comazzi M, Invernizzi G, Levinson DF, Wildenauer D, Mowry BJ, Collier D, Powell J, Crowe RR, Andreasen NC, Silverman JM, Mohs RC, Murray RM, Walters MK, Lennon DP, Hayward NK, Albus M, Lerer B, Maier W, Crow TJ (1997): A linkage study of schizophrenia to markers within Xp11 near the MAOB gene. Psychiatry Res 70:131–143.

DeLisi LE, Crow TJ (1989): Evidence for a sex chromosome locus for schizophrenia. Schizophr Bull 15:431–440.

DeLisi LE, Devoto M, Lofthouse R, Poulter M, Smith A, Shields G, Bass N, Chen G, Vita A, Morganti C, Ott J, Crow TJ (1994): Search for linkage to schizophrenia on the X and Y chromosomes. Am J Med Genet 54: 113–121.

DeLisi LE, Kelly M, Shaw S, Cardon L, Smith AB, Hopkins P, Shields G, Loftus L, Laval S, Crow TJ, Sherrington R (1997): Preliminary results of an international genomic scan for schizophrenia. Am J Med Genet 74:558. Presented at the Fifth World Congress of Psychiatric Genetics, October 20, 1997.

Dib CS, Faure S, Fizames C, Samson D, Dourot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J (1996): A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152–154.

Garner C, Kelly M, Cardon L, Joslyn G, Carey A, LeDuc C, Lichter J, Harris T, Loftus J, Shields G, Comazzi M, Vita A, Smith AM, Dann J, Crow TJ, DeLisi LE (1996): Linkage analyses of schizophrenia to chromosome 6p24–p22: An attempt to replicate. Am J Med Genet 67:595–610.

Gill M, Vallada H, Collier D, et al. (1996): A combined analysis lod D22S278 marker alleles in affected sib-pairs: Support for a susceptibility locus for schizophrenia at chromosome 22q12. Schizophrenia Collaborative Linkage Group (Chromosome 22). Am J Med Genet 1:40–45.

Gurling H, Kalsi G, Chen AH-S, Green M, Butler R, Read T, Murphy P (1995): Schizophrenia susceptibility and chromosome 6p24–22. Letter to the Editor. Nat Genet 11:234–235.

Hall JM, LeDuc CA, Watson AR, Roter AH (1996): An approach to high-throughput genotyping. Genome Res 6:781–790.

Horvatta I, Varilo T, Suvisaari J, Vaisanen JD, Terwilliger J, Lonnqvist J, Peltonen L (1997): A genome-wide search for schizophrenia genes in an internal isolate in Finland. Am J Med Genet 74:559.

Kalsi G, Brynjolfsson J, Butler R, et al. (1995): Linkage analysis of chromosome 22q12–q13 in a United Kingdom/Icelandic sample of 23 multiplex schizophrenia families. Am J Med Genet 60:298–301.

Kendler KS (1988): The genetics of schizophrenia: An overview. In Tsuang MT, Simpson JC (eds): "Handbook of Schizophrenia, Volume 3, Nosology, Epidemiology and Genetics of Schizophrenia." Amsterdam: Elsevier, pp 437–462.

Kendler KS, McGuire M, Gruenberg AM, Spellman M, O'Hare A, Walsh D (1993): The Roscommon family study. II. The risk of non-schizophrenic nonaffective psychoses in relatives. Arch Gen Psychiatry 50:645–652.

Kendler KS, Neale MC, Walsh D (1995): Evaluating the spectrum concept of schizophrenia in the Roscommon family study. Am J Psychiatry 152:749–754.

Kendler KS, MacLean CJ, O'Neill FA, Burke J, Murphy B, Duke F, Shinkwin R, Easter SM, Webb BT, Zhang J, Walsh D, Straub RE (1996): Evidence for a schizophrenia vulnerability locus on chromosome 8p in the Irish study of high-density schizophrenia families. Am J Psychiatry 153:1534–1540.

Kety SS, Wender PH, Jacobsen B, Ingraham LJ, Jansson L, Faber B, Kinney DK (1994): Mental illness in the biological and adoptive relatives of schizophrenic adoptees. Replication of the Copenhagen study in the rest of Denmark. Arch Gen Psychiatry 51:442–455.

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996): Parametric and nonparametric linkage analysis: A unified multipoint approach. Am J Hum Genet 58:1347–1363.

Lander E, Kruglyak L (1995): Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. Nat Genet 11:241–247.

Lathrop GM, Lalouel JM, Julier C, Ott J (1984): Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81:3443–3446.

Levinson DF, Mahtani MM, Brown D, Nancarrow DJ, Kirby A, Daly M, Crowe RR, Andreasen N, Sliverman JM, Mohs RC, Endicott J, Sharpe L, Walters MK, Lennon DP, Hayward NK, Kruglyak L, Mowry BJ (1996): As genome scan of schizophrenia. Presented at the 1996 World Congress on Psychiatric Genetics, San Francisco, California, October 28, 1996. Psychiatr Genet 6:140.

Lin MW, Curtis D, Williams N, et al. (1995): Suggestive evidence for linkage of schizophrenia to markers on chromosome 13q14.1–q32. Psychiatr Genet 3:117–126.

Matise TC, Perlin M, Chakravarti A (1994): Automated construction of genetic linkage maps using an expert system (MultiMap): A human genome linkage map. Nat Genet 6:384–390.

McGuffin P, Farmer AE, Gottesman II, Murray RM, Reveley AM (1984): Twin concordance for operationally defined schizophrenia. Confirmation of familiarity and heritability. Arch Gen Psychiatry 41:541–545.

Moises HW, Yang L, Kristbjarnarson H, Wiese C, Byerley W, Macciardi F, Arolt V, Blackwood D, Liu X, Sjogren B, Aschauer HN, Hwu H-G, Jang K, Livesley WJ, Kennedy JL, Zoega T, Ivarsson O, Bui M-T, Yu M-H, Havsteen B, Commenges D, Weissenbach J, Schwinger E, Gottesman II, Pakstis AJ, Wetterberg L, Kidd KK, Helgason T (1995a): An international two-stage genome-wide search for schizophrenia susceptibility genes. Nat Genet 11:321–324.

Moises HW, Yang L, Li T, Havsteen B, Fimmers R, Baur MP, Liu X, Gottesman II (1995b): Potential linkage disequilibrium between schizophrenia and locus D22S278 on the long arm of chromosome 22. Am J Med Genet 5:465–467.

Mowry BJ, Nancarrow DJ, Lennon DP, Sandkuijl LA, Crowe RR, Silverman JM, Mohs RC, Siever LJ, Endicott J, Sharpe L, Walters MK, Hayward NK, Levinson DF (1995): Schizophrenia susceptibility and chromosome 6p24–22. Letter to the Editor. Nat Genet 11:233–234.

NIMH Genetics Initiative (1997): A genome-wide search for schizophrenia susceptibility loci: The NIMH Genetics Initiative and Millennium Consortium. Presented by S. Faraone at the Fifth World Congress of Psychiatric Genetics, Santa Fe, New Mexico, October 20, 1997. Am J Med Genet 74:556–557.

Nurnberger JI, Blehar MC, Kaufman CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, Severe JB, Malaspina D, Reich T (1994): Diagnostic interview for genetic studies: Rationale, unique features and training. Arch Gen Psychiatry 51:849–862.

Ott J (1991): "Analysis of Human Linkage." Baltimore: Johns Hopkins University Press.

Parsian A, Suarez BK, Isenberg K, Hampe CL, Fisher L, Chakraverty S, Meszaros K, Lenzinger E, Willinger U, Fuchs K, Aschauer HN, Cloninger CR (1997): No evidence for a schizophrenia susceptibility gene in the vicinity of IL2RB on chromosome 22. Am J Med Genet 74:361–364.

Pfohl B, Blum N, Zimmerman M, Stangl D (1990): "Structured interview for DSM-III-R Personality Disorders-Revised (SIDP-R)." Department of Psychiatry, University of Iowa.

Polymeropoulos MH, Coon H, Byerley W, Gershon ES, Goldin L, Crow TJ, Rubenstein J, Hoff H, Holik J, Smith AM, Shields G, Bass NJ, Poulter

M, Lofthouse R, Vita A, Morganti C, Merril CR, DeLisi LE (1994): Search for a schizophrenia susceptibility locus on human chromosome 22. Am J Med Genet 54:93–99.

Polymeropoulos MH, Devoto M, Rubenstein J, Lofthouse R, Smith A, Shields G, Bass N, Crow TJ, DeLisi LE (1995): A random search of the genome in nuclear families with schizophrenia. Psychiatr Genet 5:51.

Pulver AE, Karayiorgou M, Wolyniec PS, Lasseter VK, Kasch L, Nestadt G, Antonarakis S, Housman D, Kazazian HH, Meyers D, Ott J, Lamacz M, Liang K-Y, Hanfelt J, Ullrich G, DeMarchi N, Ramu E, McHugh PR, Adler L, Thomas M, Carpenter WT, Manschreck T, Gordon CT, Kimberland M, Babb R, Puck J, Childs B (1994a): Sequential strategy to identify a susceptibility gene for schizophrenia: Report of potential linkage on chromosome 22q12–q13.1: Part I. Am J Med Genet 54:36–43.

Pulver AE, Karayiorgou M, Lasseter VK, et al. (1994b): Follow-up of a report of a potential linkage for schizophrenia on chromosome 22q12–q13.1: Part 2. Am J Med Genet 54:44–50.

Pulver AE, Lasseter VK, Kasch L, et al. (1995): Schizophrenia: A genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes. Am J Med Genet 60:252–260.

Risch N (1990): Linkage strategies for genetically complex traits. II. The power of affected relative pairs. Am J Hum Genet 46:229–241.

Sawcer S, Jones HB, Judge D, Visser F, Compston A, Goodfellow PN, Clayton D (1997): Empirical genome-wide significance levels established by whole genome simulations. Genet Epidemiol 14:223–229.

Schizophrenia Linkage Collaborative Group for Chromosomes 3, 6, and 8 (1996): Additional support for a schizophrenia linkage to chromosomes 6 and 8. Am J Med Genet 67:580–594.

Schwab SG, Albus M, Hallmayer J, et al. (1995a): Evaluation of a susceptibility gene for schizophrenia on chromosome 6P by multipoint affected sib-pair analysis. Nat Genet 11:325–327.

Schwab SG, Lerer B, Albus M, Maier W, Hallmeyer J, Fimmers R, Lichtermann D, Minges J, Bondy B, Ackenheil M, Altmark D, Hasib D, Gur E, Ebstein RP, Wildenauer DB (1995b): Potential linkage for schizophrenia on chromosome 22q12–q13: A replication study. Am J Med Genet 60:436–443.

Sherrington R, Brynjolfsson J, Petursson H, Potter M, Dudleston K, Barraclough B, Wasmuth J, Dobbs M, Gurling H (1988): Localization of a susceptibility locus for schizophrenia on chromosome 5. Nature 336: 164–167.

Silverman JM, Greenberg DA, Altstiel LD, et al. (1996): Evidence of a locus for schizophrenia and related disorders on the short arm of chromosome 5 in a large pedigree. Am J Med Genet 67:162–171.

Spitzer RL, Endicott J (1978): "Schedule for Affective Disorders and Schizophrenia (SADS)." New York: Biometrics Research Division, New York State Psychiatric Institute.

Straub RE, MacLean CJ, O'Neill FA, Burke J, Murphy B, Duke F, Shinkwin R, Webb BT, Zhang J, Walsh D, Kendler KS (1995): A potential vulnerability locus for schizophrenia on chromosome 6p24–22: Evidence for genetic heterogeneity. Nat Genet 11:287–293.

Straub RE, MacLean CJ, O'Neill FA, Walsh D, Kendler KS (1997a): Genome scan for schizophrenia genes. Presented at the Fifth World Congress of Psychiatric Genetics, Santa Fe, New Mexico, October 20, 1997. Am J Med Genet 74:558.

Straub RE, MacLean CJ, O'Neill FA, Walsh D, Kendler KS (1997b): Support for a possible schizophrenia vulnerability locus in a region of 5q22–31 in Irish families. Mol Psychiatry 2:148–155.

Utah Marker Development Group (1995): A collection of ordered tetranucleotide-repeat markers from the human genome. Am J Hum Genet 57:619–628.

Wang S, Sun C, Walczak CA, Ziegle JS, Kipps BR, Goldin LR, Diehl SR (1995): Evidence for a susceptibility locus for schizophrenia on chromosome 6pter–p22. Nat Genet 10:41–46.

Wildenauer DN (1996): Chromosome 18 and schizophrenia. Presented at the 1996 Annual Meeting of the ACNP, San Juan, Puerto Rico, December, 1996.

Wildenauer DB, Albus M, Schwab SG, Hallmayer J, Hanses C, Eckstein GN, Zill P, Honig S, Lerer B, Ebstein R, Lichtermann D, Trixler M, Borrmann M, Maier W (1997): Searching for susceptibility genes in schizophrenia by affected sib-pair analysis. Presented at the Fifth World Congress of Psychiatric Genetics, Santa Fe, New Mexico, October 20, 1997. Am J Med Genet 74:558.

Williams NM, Rees MI, Holmans P, Daniels J, Fenton I, Cardno AG, Murphy KC, Jones LA, Asherson P, McGuffin P, Owen MJ (1997): Genome search for schizophrenia susceptibility genes using a two-stage sib-pair approach. Presented at the Fifth World Congress of Psychiatric Genetics, Santa Fe, New Mexico, October 20, 1997. Am J Med Genet 74:559.

Schizophrenia Collaborative Linkage Group (Chromosome 22) (1996): A combined analysis of D22S278 marker alleles in affected sib-pairs: Support for a susceptibility locus for schizophrenia at chromosome 22q12. Am J Med Genet 67:40–45.

Vallada HP, Gill M, Sham P, Lim LC, Nanko S, Asherson P, Murray RM, McGuffin P, Owen M, Collier D (1995): Linkage studies on chromosome 22 in familial schizophrenia. Am J Med Genet 60:139–146.

Exhibit N

Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 5604–5609, May 1999
Genetics

A high-density genome scan detects evidence for a bipolar-disorder susceptibility locus on 13q32 and other potential loci on 1q32 and 18p11.2

SEVILLA D. DETERA-WADLEIGH*†, JUDITH A. BADNER*‡, WADE H. BERRETTINI§, TAKEO YOSHIKAWA*, LYNN R. GOLDIN*¶, GORDON TURNER*, DENISE Y. ROLLINS*, TRACY MOSES*, ALAN R. SANDERS*‡, JAYAPRAKASH D. KARKERA*, LISA E. ESTERLING*, JIN ZENG*, THOMAS N. FERRARO§, JULIET J. GUROFF*, DIANE KAZUBA*, MARY E. MAXWELL*, JOHN I. NURNBERGER, JR.‖, AND ELLIOT S. GERSHON*‡

*Clinical Neurogenetics Branch, National Institute of Mental Health, and ¶Genetic Epidemiology Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892; ‡Department of Psychiatry, University of Chicago, Chicago, IL 60637; §Center for Neurobiology and Behavior, Department of Psychiatry, University of Pennsylvania, Philadelphia, PA 19104; and ‖Medical and Molecular Genetics and Psychiatry, Indiana University, Indianapolis, IN 46202

Edited by Newton E. Morton, University of Southampton, United Kingdom, and approved March 3, 1999 (received for review December 18, 1998)

ABSTRACT **Bipolar disorder is a severe mental illness characterized by mood swings of elation and depression. Family, twin, and adoption studies suggest a complex genetic etiology that may involve multiple susceptibility genes and an environmental component. To identify chromosomal loci contributing to vulnerability, we have conducted a genome-wide scan on ≈396 individuals from 22 multiplex pedigrees by using 607 microsatellite markers. Multipoint nonparametric analysis detected the strongest evidence for linkage at 13q32 with a maximal logarithm of odds (lod) score of 3.5 (P = 0.000028) under a phenotype model that included bipolar I, bipolar II with major depression, schizoaffective disorder, and recurrent unipolar disorder. Suggestive linkage was found on 1q31-q32 (lod = 2.67; P = 0.00022) and 18p11.2 (lod = 2.32; P = 0.00054). Recent reports have linked schizophrenia to 13q32 and 18p11.2. Our genome scan identified other interesting regions, 7q31 (lod = 2.08; P = 0.00099) and 22q11-q13 (lod = 2.1; P = 0.00094), and also confirmed reported linkages on 4p16, 12q23-q24, and 21q22. By comprehensive screening of the entire genome, we detected unreported loci for bipolar disorder, found support for proposed linkages, and gained evidence for the overlap of susceptibility regions for bipolar disorder and schizophrenia.**

Bipolar disorder is a complex neuropsychiatric disease of unknown genetic etiology. Theories concerning multigenic inheritance involving multiple genes that exert varying magnitudes of effect on overall susceptibility have been advanced, but these genes have yet to be identified. Early linkage studies on bipolar disorder have been marked with uncertainty (1), possibly because of difficulties related to the power of detecting weak- to moderate-effect genes (2–4), to other confounding factors, or to spurious reported findings. Recent advances, which include the availability of dense maps containing highly informative markers, the use of extended pedigrees or larger sample collections, and the application of newer analytical methods, have propelled reports showing evidence of linkage to chromosomes 4p16 (5), 12q23-q24 (6), 18pcen-qcen (7, 8), 18q22-q23 (9), and 21q22.3 (10).

Previous chromosomal screens on our 22 multiply affected pedigrees elicited evidence of linkage to chromosome 18 (7, 8) and support for the proposed linkage on 21q22.3 (11). In contrast, our earlier scans that used sparsely spaced and moderately informative markers have not uncovered areas of linkage (12, 13). In the present study, we performed a high-density genome scan by using highly polymorphic markers to identify chromosomal regions that contribute to the genetic risk of bipolar disorder. Here, we report evidence for bipolar-disorder susceptibility loci, support of prior linkages, and possible overlaps of regions implicated in bipolar disorder and schizophrenia.

METHODS

Pedigrees. A panel of 22 multiplex pedigrees was used for the genome screening, and the diagnostic and ascertainment methods used for 21 of these have been described in detail elsewhere (14). These families included ≈365 informative persons (i.e., those persons whose genotype can be determined either directly or indirectly; ref. 14). The 22nd family was the "right extension" of the Old Order Amish pedigree 110 (7).

To address the issue of uncertainty regarding the spectrum of illness that is inherited, two hierarchical models of the affected phenotype were used in linkage analysis. For affection status model (ASM) I, the definition of the affected category was restricted to bipolar I, bipolar II with major depression, and schizoaffective disorder. For the second model, ASM II, the phenotype definition was broadened to include those individuals with two or more episodes of major depression. Each pedigree had a minimum of four affected individuals under ASM II. There were 396 individuals that were available for genotyping, including 117 with ASM I and 159 with ASM II diagnoses.

Although we excluded branches of a pedigree when a person marrying into the pedigree had any of the diagnoses included in ASM I and ASM II, bilineality could not be ruled out, particularly if a multigenic model with common susceptibility alleles is assumed. However, when nonparametric methods are used, this issue may not be relevant. Bilineal families may reduce power but would not be expected to increase the incidence of false-positive results.

The power of this pedigree series to detect linkage has been examined in a prior study; this study used assumed parameters, because the true genetic model for bipolar disorder is not known (14). Simulations of linkage using 21 of the 22 families included in the present genome scan were performed under a dominant transmission model and ASM II by using the SIMLINK program (15). Assuming the presence of heterogeneity, there

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

This paper was submitted directly (Track II) to the *Proceedings* office. Abbreviations: lod, logarithm of odds; ASM, affection status model; ASP, affected sib pair; GHP, the GENEHUNTER-PLUS program; GHPfam, GHP analysis with pedigrees split into nuclear families; cM, centimorgan; IBD, identical by descent.

†To whom reprint requests should be addressed. e-mail: sevilla@codon.nih.gov.

was >50% power when at least 25% of the families are linked, for a marker with a polymorphism information content of 0.7 at $\Theta = 0.01$. In a multigenic model, power depends on the effect of the susceptibility gene and whether it is acting in a multiplicative or additive fashion with other predisposing genes.

Genotyping. The microsatellite markers used for genotyping consisted of tandem repeated dinucleotides and tetranucleotides, as well as a few trinucleotides. Markers were chosen from the following databases: the Cooperative Human Linkage Center (CHLC; www.chlc.org); the CHLC/Weber screening set, versions 6 and 8 (Research Genetics, Huntsville, AL); Genethon (ftp://ftp.genethon.fr/pub/Gmap/Nature-1995/); the Center for Medical Research of the Marshfield Medical Research Foundation (www.marshmed.org/genetics/); and the Genetic Location Database (ref. 16; cedar.genetics.soton.ac.uk/public_html/). A few markers were taken from the map created by the National Institutes of Health/Centre d'Étude du Polymorphisme Humain (Paris) Collaborative Mapping Group (mapper.wustl.edu/basemaps/index.html). The Marshfield map included almost all of the markers, and, in other instances, intermarker distance and order were deduced from the pedigrees by using CRIMAP [Green, P., Falls, K. & Crooks, S. (1990) CRIMAP Documentation (Department of Genetics, School of Medicine, Washington University, St. Louis), Version 2.4; linkage.rockefeller.edu/soft/crimap/].

With few exceptions, PCR amplification was done by using one set of cycling conditions: 95°C for 5 min for the initial denaturation, 30 cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 1 min. A final extension of 72°C for 10 min was done. Markers that failed under these conditions were replaced by other markers mapping to the same region. The controls CEPH 1333-01 and 1333-02 were placed in specific lanes with each set of samples to obtain consistent assignment of allele sizes from gel to gel. Most of the genotypings were radioactivity-based, and each autoradiogram was read by two independent readers who were blind to diagnosis. As size standards, we used an M13 sequence ladder and a 123-bp ladder. Automated genotyping was done with an Applied Biosystems (ABI) 373 sequencer (Perkin–Elmer), and alleles were scored with the GENESCAN and GENOTYPER programs. Unlabeled and fluorescently labeled primers were purchased from Research Genetics, although a few were synthesized through BioServe Biotechnology (Gaithersburg, MD). Mendelian inconsistencies were resolved before linkage analysis, and marker allele frequencies were computed from individuals marrying into the pedigree.

Linkage Analysis. Linkage was evaluated under the affected phenotype classifications of ASM I and ASM II. Pointwise affected-sib-pair (ASP) analysis was done by including all ASPs by using the SIBPAL program, version 3.0, in the SAGE package (17). Multipoint analyses were performed by using ASPEX [Hinds, D. & Risch, N. (1996) ASPEX, version 1.62, sib_phase package; ftp://lahmed.stanford.edu/pub/aspex] and GENEHUNTER-PLUS (GHP), version 1.1 (18, 19). Sex-averaged maps were used for multipoint analyses. For chromosome 18, sex-specific maps were used in additional calculations. For the GHP analysis, individuals that were uninformative or that were connected to the pedigree by more than two uninformative generations were deleted. The right extension of the Amish pedigree 110 was divided into two families. Pedigrees were analyzed using the ALL function of GHP, which examines all individuals simultaneously and assigns a higher score when more of them share the same allele by descent. The probability of the observed sharing is calculated by using a semiparametric logarithm of odds (lod) score as described (19).

GHP analysis also was done after splitting the pedigrees into nuclear families (GHPfam). It has been shown through simulation that in GHP analysis, more distant relatives provide less information when a susceptibility allele is common (20). The reverse is true when the allele is rare. GHPfam was used when there was a significant difference between GHP and ASPEX data. If the results from GHPfam are similar to GHP, then the differences likely were caused by the different algorithms. On the other hand, if the GHPfam results are similar to ASPEX, then the family structure being analyzed likely is affecting the results.

Parametric lod score analysis was conducted by using FASTLINK, version 3.0 P [Schaffer, A. A., Gupta, S. K., Shriram, K. & Cottingham, R. W., Jr. (1997); ftp://fastlink.nih.gov/pub/fastlink]. Pairwise lod score calculations were done under a dominant model with 85% and 50% penetrance and under a recessive model with 85% penetrance.

Because we have employed different methods of analysis under two phenotype definitions, the most significant signals in any specific region may be inflated because of multiple testing. Biases in each analysis caused by deviations from assumptions also could occur. We believe, however, that the consistency of results derived from different analytical methods is important to consider, particularly if two or more methods detect lod scores of >2 in the same region. Moreover, support from other studies on independently ascertained families is critical to strengthening the validity of a finding.

RESULTS AND DISCUSSION

Genome-Wide Screening Data Analyzed by Nonparametric Methods. We have screened 607 polymorphic loci covering all 22 autosomes and the X chromosome, with an average marker spacing of ≈6 centimorgans (cM) and an average heterozygosity of ≈60%. Model-free calculations identified clusters of loci in several regions of the genome that displayed excess identical-by-descent (IBD) sharing.

The strongest evidence for linkage derived from nonparametric analysis centered at chromosome 13q32, the only region that yielded a lod score of >3 (Table 1; Fig. 1). Pointwise analysis depicted excess IBD sharing in an ≈17-cM region maximizing at D13S1271 (64%; $P = 0.0002$) under ASM I. Analysis by GHPfam using nuclear families detected the highest multipoint lod at the D13S1252–D13S1271 interval (lod = 3.5; $P = 0.000028$) under ASM II. ASPEX localized the maximal peak in the D13S1271–D13S779 region for both ASM I (lod = 3.4; $P = 0.000039$) and ASM II (lod = 3.3; $P = 0.000051$). Broadening the classification of affected phenotypes to include recurrent major depression did not have a significant effect on the lod score. Based on statistical criteria proposed by Lander and Kruglyak (21), these values approach significant linkage (lod = 3.6; $P = 0.000022$) for a genome scan. Consistent with this finding was an excess sharing in this region in the 97 bipolar

Table 1. Multipoint linkage data derived through ASPEX showing lod of >1 ($P \leq 0.01$)

Chromosome	Region	ASM I (*P*)	ASM II (*P*)
1q25-q32	S471–S237	1.78 (0.0021)*	
2pter	S2976		2.0 (0.0012)
2p25-p24	S1400–S1360	1.11 (0.012)	
4p16-p14	S2408–S2632		1.77 (0.0022)
5q33-q35	S498–S408	1.16 (0.01)	1.7 (0.0026)
11q13	S1883–S913	1.89 (0.0016)	
12q22-q24	S1343–S2070		1.24 (0.0084)
13q32	S1271–S779	3.4 (0.000039)	3.3 (0.000051)†
14q24	S1434–S65	1.04 (0.014)	
14q24-q32.2	S617–S1434		1.23 (0.0087)
18p11.2	S1150–S71	2.32 (0.00054)	2.03 (0.0011)
21q22.1	S1254–S65	1.85 (0.0018)	
22q11-q13	S689–S685	1.26 (0.008)	2.1 (0.00094)

*GHPfam yielded lod = 2.67 ($P = 0.00022$) at D1S1660–D1S1678.
†GHPfam yielded lod = 3.5 ($P = 0.000028$) at D13S1252–D13S1271.

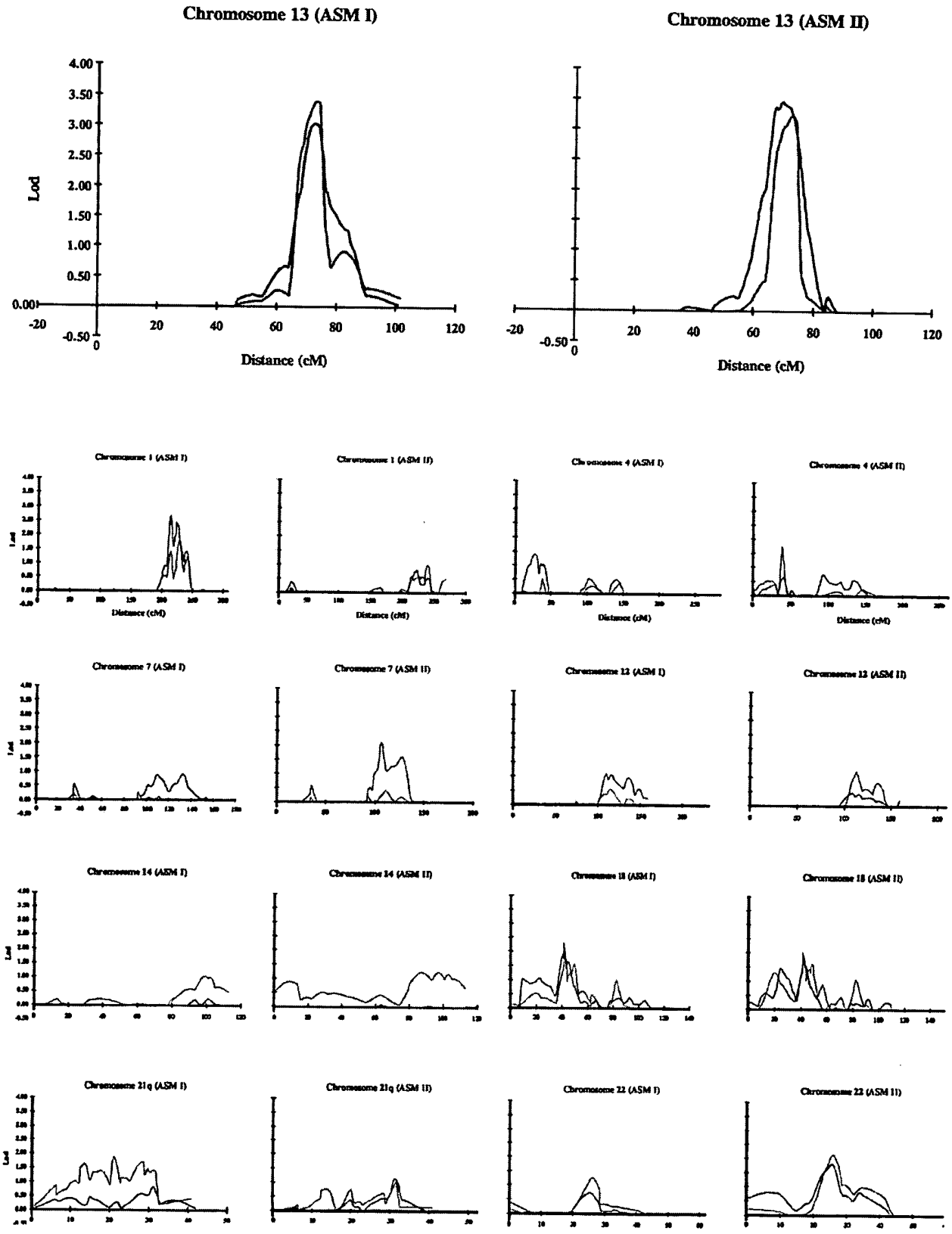


FIG. 1. Multipoint nonparametric analysis of a genome scan on chromosomes 13, 1, 4, 7, 12, 14, 18, 21, and 22. The affected phenotype models ASM I and ASM II are defined in the text. Plots were derived through ASPEX (red), GHPfam (purple), and GHP (blue).

pedigrees of the National Institute of Mental Health Genetics Initiative (22). These data suggest a prominent role for a gene encoded by 13q32 in the genetic basis of affective disorder in these pedigrees. It is noteworthy that the susceptibility region

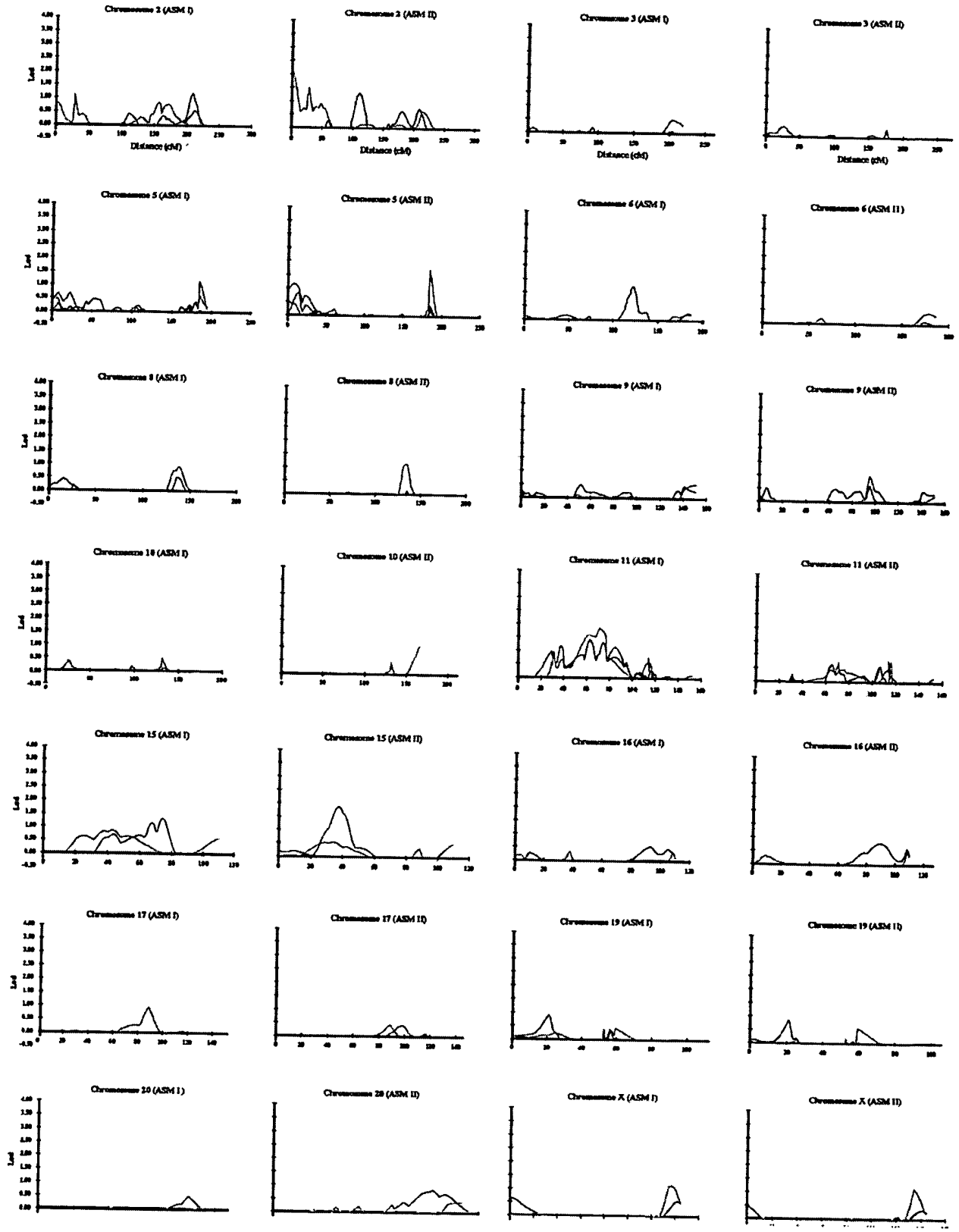


FIG. 2. Multipoint nonparametric analysis of a genome scan on chromosomes 2, 3, 5, 6, 8, 9, 10, 11, 15, 16, 17, 19, 20, and X. ASM I and ASM II are defined in the text. Analysis was done with ASPEX (red) and GHP (blue).

on 13q32 overlapped the most significant region in a schizophrenia series (23). Examination of the diagnosis in the current pedigrees showed that there was at least one individual with schizoaffective disorder or schizophrenia in 15 of the families.

Conceivably, the 13q32 gene is a predisposing factor for psychoses.

Regions on chromosomes 1, 7, 18, and 22 showed excessive allele sharing, yielding maximal multipoint lods of >2 but <3 (Table 1; Fig. 2), with values for 1q and 18p exceeding suggestive linkage (21). Pointwise analysis disclosed elevated IBD sharing, in an ≈30-cM region spanning 1q25-q42, which was more pronounced under ASM I. Addition of recurrent major depression to the stringent phenotypes, as in ASM II, may increase power, because it expands the number included in the affected category. On the contrary, this addition could have the opposite effect, because the genetic heterogeneity of the sample would be magnified. Multipoint analysis by GHPfam produced a peak of allele sharing at 1q25-q32 (ASM I lod = 2.67; $P = 0.00022$). Complementing this finding was a study that showed a frequent fragile site on 1q32 in bipolar patients (24). Together, these results raise the possibility that disruption of a gene on 1q32 contributes to overall susceptibility to bipolar disorder.

The candidate region on chromosome 18 encompassed a large area surrounding the centromere (7, 8). We increased marker coverage to an average spacing of 3.2 cM and localized the highest peak between D18S1150 and D18S71 (ASM I lod = 2.32; $P = 0.00054$) at 18p11.2 (25) by using ASPEX (Fig. 1). There was no significant discrepancy between the peak profiles for ASM I and ASM II. A telomeric, less prominent peak at ≈18p11.3 (ASM II GHP lod = 1.44; $P = 0.005$) and another small peak at ≈18q21 (lod = ≈1) also were noted. These weak signals may be relevant because an inversion involving 18p11.3 and 18q21.1 seemed to cosegregate with affective disorder in some families (26), and 18q21 included a proposed region for bipolar disorder (27, 28).

Additional multipoint ASPEX analysis was performed by using both male- and female-specific maps (data not shown). Both analyses generated peak height and location that were roughly similar to those found when the sex-averaged map was used (Fig. 1). A previous study postulated a parent-of-origin effect on bipolar disorder in this chromosome (27). We extended the analysis to calculate parental-specific lods by using sex-specific maps. Paternal allele sharing under ASM I was increased, consistent with a previous finding on these pedigrees (29). However, this increase seemed to be caused by a single large family in which the paternal haplotype was shared (data not shown).

A recent study implicated 18p11.2 in psychoses (30). Evidence for linkage and association in schizophrenia families increased when affective diagnoses were included in the affected category (30). The emergence of overlaps for bipolar disorder and schizophrenia loci on 13q32 (23) and 18p11.2 (30) is consistent with family studies that showed an increased risk among first degree relatives of schizophrenia probands for schizoaffective and recurrent unipolar disorders (31, 32).

Neither SIBPAL nor ASPEX detected a region of excess IBD sharing on chromosome 7. In contrast, GHP gave a lod of 2.08 ($P = 0.00099$) under ASM II between D7S1799 and D7S501 at 7q31. Coincident with this location was an area with elevated allele sharing in the bipolar series of the National Institute of Mental Health Genetics Initiative (33). The divergent results derived through ASPEX and GHP may be ascribed to random chance or to a rare susceptibility allele that is more likely to be detected when extended pedigrees are studied (20).

On chromosome 22, the maximal signal spanned D22S689 and D22S685 on 22q11-q13 (ASM II ASPEX lod = 2.1; $P = 0.00094$). In this region, microdeletions have been found in velo-cardio-facial syndrome, a dysmorphism associated with an enhanced prevalence of psychoses (34). Significant linkage to bipolar disorder (J. Kelsoe, personal communication) and support for a susceptibility locus in schizophrenia (23, 35) have been reported for this region of chromosome 22.

We have generated evidence supporting proposed linkages (21) on chromosomes 4p16 (see below), 12q23-q24, and 21q22 (Table 1; Fig. 1). The 12q23-q24 region has been linked to Darier's disease (36), which was found to cosegregate with affective disorder (37). Screening of loci in the region identified significant linkage (21) to bipolar disorder in Danish kindreds (6) and in an extended French-Canadian pedigree (N. Barden, personal communication). Confirmatory data were detected in our families (ASM II ASPEX lod = 1.24; $P = 0.0084$). We have extended our earlier study on 21q22 (11) and found a broad region with excess sharing that included a proposed vulnerability locus (10), maximizing at D21S1254–D21S65 (ASM I ASPEX lod = 1.85; $P = 0.0018$).

Nonparametric analysis detected other loci with marginal signals (lod of >1 but <2; Table 1; Fig. 2). Excess IBD sharing (lod = 1.77) was found on 5q33-q35 (Table 1), and by two point analysis under a dominant transmission (85% penetrance), D5S462 gave a lod of 1.8 (Table 2). Lod scores of >1 under a dominant model have been reported for markers mapping to this region (38). Excess sharing was displayed by a marker mapping to 5q35 in the National Institute of Mental Health Genetics Initiative bipolar pedigrees (39). The increase in IBD sharing on chromosome 11q13 (lod = ≈1.9) seemed to be supported by parametric analysis, because D11S4076, located in this region, gave a lod of 1.55 under a dominant transmission (Table 2). A cautious interpretation of these signals is warranted, because several methods of analysis under two phenotype models were employed; therefore, signals could have arisen from random statistical fluctuations.

Genome-Wide Screening Data Analyzed by a Parametric Method. Parametric calculations identified two loci giving lods

Table 2. Pairwise parametric lod scores >1.5 ($\theta = 0.2$) for the entire pedigree series

Chromosome	Locus	ASM I	ASM II
1q32	GATA124F08	2.37 (dom, 85% pen)	
5q35	D5S462	1.80 (dom, 85% pen)	1.54 (dom, 85% pen)
7p15-q22	D7S492		1.59 (dom, 50% pen)
10q25	D10S187		1.69 (rec, 85% pen)
11p15-p14	D11S915	1.84 (dom, 50% pen)	
11p14-p13	D11S904	1.62 (dom, 50% pen)	
11q12.3-q13.2	D11S4076	1.55 (dom, 50% pen)	
13q32	D13S1271	2.06 (rec, 85% pen)	1.69 (rec, 85% pen)
14q11-q13	D14S1060		2.19 (rec, 85% pen)
14q32.1-q32.2	D14S1434	1.79 (rec, 85% pen)	
18p11.2	D18S1353	1.78 (dom, 85% pen)	
18p11.2	D18S40		1.70 (dom, 50% pen)
21q22.1-p22.3	D21S267	1.57 (rec, 85% pen)	
21q22.3-qter	D21S212	1.79 (rec, 85% pen)	

Parametric analysis was done with FASTLINK (see *Methods*). dom, dominant; rec, recessive, pen, penetrance.

of >3 in the two largest pedigrees. Pedigree 0048 had 39 members that were genotyped, with six affected under ASM I (14). On 4p16, a region that showed linkage to bipolar disorder in Scottish families (5), D4S2632 gave a lod of 3.24 ($\Theta = 0$) for ASM I under a dominant, 50% penetrance model. Model-free analysis by ASPEX showed elevated allele sharing within the D4S2408–D4S2632 interval (ASM II lod = 1.77; $P = 0.0022$; Table 1; Fig. 2). Hence, these data support the reported linkage (5), and the gene on 4p16 may confer a major effect on susceptibility in pedigree 0048.

Pedigree 1482 had 30 members genotyped with 11 affected under ASM II (14). At D14S617 on 14q23-q32.1, this pedigree gave a lod of 3.19 ($\Theta = 0$; dominant transmission; 85% penetrance) under ASM II. Approximately 8 cM distal to D14S617 on 14q32.1-q32.2, D14S1434 displayed a total lod score of 1.79 for the entire pedigree series under ASM I and a recessive, 85% penetrance model (Table 2). ASPEX analysis displayed increased sharing in the region (ASM II lod = 1.23; $P = 0.0087$; Table 1; Fig. 1). At a more proximal location, D14S1060, which maps to 14q11-q13, a region involved in a balanced translocation in a patient with bipolar disorder (40), displayed a lod of >2 under ASM II and a recessive transmission (Table 2). It is unclear whether these signals would be supported on further scrutiny of other pedigree series.

Markers that produced parametric lods of >1.5 for the entire panel tended to reinforce findings from nonparametric analysis on 1q32, 13q32, 18p11.2, and 21q22 (Table 2), even though parameters involved in the genetic inheritance of bipolar disorder are not known. GATA124F08 on 1q32 yielded a lod of >2, assuming a dominant mode of transmission, and D13S1271 on 13q32 yielded a lod of >2 under a recessive model. Similarly, the peak heterogeneity lod in a schizophrenia series for a marker on 13q32 was found by using a recessive model (23). The highest parametric lods and IBD sharing on chromosome 18 were found in loci mapping to 18p11.2 (Fig. 1). The distal portion of 21q also had markers with lods of >1.5 under a recessive model, which is consistent with findings from model-free analysis. Despite the concurrence of signals derived from different analytical methods and some apparent corroborating evidence from other studies, verification of linkage requires additional confirmation from other pedigree series and must await cloning of genes manifesting variants that cosegregate with the affected phenotype.

By implementing a high-density genome scan, we detected potential susceptibility loci for bipolar disorder on 13q32, 1q32, and 18p11.2 and found support for proposed predisposing regions. Our data, coupled with independent findings, suggest the existence of shared loci for bipolar disorder and schizophrenia and may provide evidence for the continuum model of psychosis (41).

We thank the patients and their family members for their participation in the study.

1. Risch, N. & Botstein, D. (1996) *Nat. Genet.* **12**, 351–353.
2. Hauser, E. R., Boehnke, M., Guo, S.-W. & Risch, N. (1996) *Genet. Epidemiol.* **13**, 117–137.
3. Suarez, B. K., Hampe, C. L. & Van Eerdewegh, P. (1994) in *Genetic Approaches to Mental Disorders*, eds. Gershon, E. S. & Cloninger, C. R. (American Psychiatric Institute, Washington, DC), pp. 23–46.
4. Risch, N. & Merikangas, K. (1996) *Science* **273**, 1516–1517.
5. Blackwood, D. H. R., He, L., Morris, S. W., McLean, A., Whitton, C., Thomson, M., Walker, M. T., Woodburn, K., Sharp, C. M., Wright, A. F., *et al.* (1996) *Nat. Genet.* **12**, 427–430.
6. Ewald, H., Degn, B., Mors, O. & Kruse, T. A. (1998) *Psychiatr. Genet.* **8**, 131–140.
7. Berrettini, W. H., Ferraro, T. N., Goldin, L. R., Weeks, D. E., Detera-Wadleigh, S. D., Nurnberger, J. I., Jr., & Gershon, E. S. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 5918–5921.
8. Berrettini, W. H., Ferraro, T. N., Choi, H., Goldin, L. R., Detera-Wadleigh, S. D., Muniec, D., Hsieh, W.-T., Hoehe, M., Guroff, J. J., Kazuba, D., *et al.* (1997) *Arch. Gen. Psychiatry* **54**, 32–39.
9. Freimer, N. B., Reus, V. I., Escamilla, M. E., McInnes, L. A., Spesny, M., Leon, P., Service, S. K., Smith, L. B., Silva, S., Rojas, E., *et al.* (1996) *Nat. Genet.* **12**, 436–441.
10. Straub, R. E., Lehner, T., Luo, Y., Loth, J. E., Shao, W., Sharpe, L., Alexander, J. R., Das, K., Simon, R., Fieve, R. R., *et al.* (1994) *Nat. Genet.* **8**, 291–296.
11. Detera-Wadleigh, S. D., Badner, J. A., Goldin, L. R., Berrettini, W. H., Sanders, A. R., Rollins, D. Y., Turner, G., Moses, T., Hacrian, H., Muniec, D., *et al.* (1998) *Am. J. Hum. Genet.* **58**, 1279–1285.
12. Berrettini, W. H., Detera-Wadleigh, S. D., Goldin, L. R., Martinez, M., Hsieh, W.-T, Hoehe, M., Choi, H., Muniec, D., Ferraro, T. N., Guroff, J. J., *et al.* (1991) *Psychiatr. Genet.* **2**, 191–208.
13. Detera-Wadleigh, S. D., Hsieh, W.-T., Berrettini, W. H., Goldin, L. R., Rollins, D. Y., Muniec, D., Grewal, R., Guroff, J. J., Turner, G., Coffman, D., *et al.* (1994) *Am. J. Med. Genet.* **54**, 206–218.
14. Berrettini, W. H., Goldin, L. R., Martinez, M. M., Maxwell, E., Smith, A. L., Guroff, J. J., Kazuba, D. M., Nurnberger, J. I., Jr., Hamovit, J., Simmons-Alling, S., *et al.* (1991) *Psychiatr. Genet.* **2**, 125–160.
15. Boehnke, M. (1986) *Am. J. Hum. Genet.* **39**, 513–527.
16. Collins, A., Frezal, J., Teague, J. & Morton, N. E. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 14771–14775.
17. Department of Epidemiology and Biostatistics (1997) SAGE, Statistical Analysis for Genetic Epidemiology (Case Western Reserve Univ., Cleveland), Release 3.0.
18. Kruglyak, L., Daly, M. J., Reeve-Daly, M. P. & Lander, E. S. (1996) *Am. J. Hum. Genet.* **58**, 1347–1363.
19. Kong, A. & Cox, N. J. (1997) *Am. J. Hum. Genet.* **61**, 1179–1188.
20. Badner, J. A., Gershon, E. S. & Goldin, L. R. (1998) *Am. J. Hum. Genet.* **63**, 880–888.
21. Lander, E. S. & Kruglyak, L. (1995) *Nat. Genet.* **11**, 241–247.
22. Stine, O. C., McMahon, F. J., Chen, L., Xu, J., Meyers, D. A., MacKinnon, D. F., Simpson, S., McInnis, M. G., Rice, J. P., Goate, A., *et al.* (1997) *Am. J. Med. Genet.* **74**, 263–269.
23. Blouin, J.-L., Dombroski, B. A., Nath, S. K., Lasseter, V. K., Wolyniec, P. S., Nestadt, G., Thornquist, M., Ullrich, G., McGrath, J., Kasch, L., *et al.* (1998) *Nat. Genet.* **20**, 70–73.
24. Turecki, G., Smith, M. & Mari, J. J. (1995) *Am. J. Med. Genet.* **60**, 179–182.
25. Esterling, L. E., Matise, T. C., Sanders, A. R., Yoshikawa, T., Overhauser, J., Gershon, E. S., Moskowitz, M. & Detera-Wadleigh, S. D. (1997) *Mol. Psychiatry* **2**, 501–504.
26. Mors, O., Ewald, H., Blackwood, D. & Muir, W. (1997) *Br. J. Psychiatry* **170**, 278–280.
27. Stine, O. C., Xu, J., Koskela, R., McMahon, F., Gschwend, M., Friddle, C., Clark, C. D., McInnis, M. G., Simpson, S. G., Breschel, T. S., *et al.* (1995) *Am. J. Hum. Genet.* **57**, 1384–1394.
28. McMahon, F. J., Hopkins, P. J., Xu, J., McInnis, M. G., Shaw, S., Cardon, L., Simpson, S. G., MacKinnon, D. F., Stine, O. C., Sherrington, R., *et al.* (1997) *Am. J. Hum. Genet.* **61**, 1397–1404.
29. Gershon, E. S., Badner, J. A., Detera-Wadleigh, S. D., Ferraro, T. N. & Berrettini, W. H. (1996) *Am. J. Med. Genet.* **67**, 202–207.
30. Schwab, S. G., Hallmayer, J., Lerer, B., Albus, M., Borrmann, M., Hönig, S., Strauss, M., Segman, R., Lichtermann, D., Knapp, M., *et al.* (1998) *Am. J. Hum. Genet.* **63**, 1139–1152.
31. Gershon, E. S., DeLisi, L. E., Hamovit, J., Nurnberger, J. I., Jr., Maxwell, M. E., Schreiber, J., Dauphinais, D., Dingman, C. W., II, & Guroff, J. J. (1988) *Arch. Gen. Psychiatry* **45**, 328–336.
32. Maier, W., Lichtermann, D., Minges, J., Hallmayer, J., Heun, R., Benkert, O. & Levinson, D. (1993) *Arch. Gen. Psychiatry* **50**, 871–883.
33. Detera-Wadleigh, S. D., Badner, J. A., Yoshikawa, T., Sanders, A. R., Goldin, L. R., Turner, G., Rollins, D. Y., Moses, T., Guroff, J. J., Kazuba, D., *et al.* (1997) *Am. J. Med. Genet.* **74**, 254–262.
34. Pulver, A. E., Nestadt, G., Goldberg, R., Shprintzen, R. J., Lamacz, M., Wolyniec, P. S., Morrow, B., Karayiorgou, M., Antonarakis, S. E. & Housman, D. (1994) *J. Nerv. Ment. Dis.* **182**, 476–478.
35. Schizophrenia Collaborative Linkage Group (Chromosome 22) (1996) *Am. J. Med. Genet.* **67**, 40–45.
36. Craddock, N., Dawson, E., Burge, S., Parfitt, L., Mant, B., Roberts, Q., Daniels, J., Gill, M., McGuffin, P., Powell, J., *et al.* (1993) *Hum. Mol. Genet.* **2**, 1941–1943.
37. Craddock, N., Owen, M., Burge, S., Kurina, B., Thomas, P. & McGuffin, P. (1994) *Br. J. Psychiatry* **164**, 355–358.
38. Coon, H., Jensen, S., Hoff, M., Holik, J., Plaetke, R., Reimherr, F., Wender, P., Leppert, M. & Byerley, W. (1993) *Am. J. Hum. Genet.* **52**, 1234–1249.
39. Edenberg, H. J., Foroud, T., Connelly, P. M., Sorbel, J. J., Carr, K., Crose, C., Willig, C., Zhao, J., Miller, M., Bowman, E., *et al.* (1997) *Am. J. Med. Genet.* **74**, 238–246.
40. Overhauser, J., Berrettini, W. H. & Rojas, K. (1998) *Psychiatr. Genet.* **8**, 53–56.
41. Crow, T. J. (1986) *Br. J. Psychiatry* **149**, 419–429.

Exhibit O

A genome survey indicates a possible susceptibility locus for bipolar disorder on chromosome 22

John R. Kelsoe*†, M. Anne Spence‡, Erika Loetscher§, Montserrat Foguet§¶, A. Dessa Sadovnick‖, Ronald A. Remick**, Pamela Flodman‡, Jason Khristich*, Zofi Mroczkowski-Parker*, John L. Brown*, Diane Masser‡, Sharon Ungerleider‡, Mark H. Rapaport*, William L. Wishart§, and Hermann Luebbert§¶

*Department of Psychiatry, University of California at San Diego, La Jolla, CA 92093, and San Diego Veterans Affairs Healthcare System, La Jolla, CA 92161; ‡Department of Pediatrics, University of California, Irvine, CA 92613; §Nervous System, Novartis Pharma AG, 4002 Basel, Switzerland; ‖Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada V6T 1Z3; and **Department of Psychiatry, St. Paul's Hospital, Vancouver, BC, Canada V6Z 1Y6

Edited by Newton E. Morton, University of Southampton, Southampton, United Kingdom, and approved November 6, 2000 (received for review July 31, 2000)

Bipolar disorder or manic depressive illness is a major psychiatric disorder that is characterized by fluctuation between two abnormal mood states. Mania is accompanied by symptoms of euphoria, irritability, or excitation, whereas depression is associated with low mood and decreased motivation and energy. The etiology is currently unknown; however, numerous family, twin, and adoption studies have argued for a substantial genetic contribution. We have conducted a genome survey of bipolar disorder using 443 microsatellite markers in a set of 20 families from the general North American population to identify possible susceptibility loci. A maximum logarithm of odds score of 3.8 was obtained at D22S278 on 22q. Positive scores were found spanning a region of nearly 32 centimorgans (cM) on 22q, with a possible secondary peak at D22S419. Six other chromosomal regions yielded suggestive evidence for linkage: 3p21, 3q27, 5p15, 10q, 13q31-q34, and 21q22. The regions on 22q, 13q, and 10q have been implicated in studies of schizophrenia, suggesting the possible presence of susceptibility genes common to both disorders.

Bipolar disorder is a severe psychiatric disorder that affects approximately 1% of the world's population (1). It is characterized by extreme swings in mood between mania and depression. Mania is accompanied by euphoria, grandiosity, increased energy, decreased need for sleep, rapid speech, and risk taking. Depression is associated with low mood, low energy and motivation, insomnia, and feelings of worthlessness and hopelessness. Psychosis can occur in either state, and there is a 17% lifetime risk for suicide.

The etiology is currently unknown, but epidemiological studies argue for a strong genetic component. Family studies indicate an approximately 7-fold increase in risk to first-degree family members (2). Twin studies find an average 4-fold increase in risk to monozygotic vs. dizygotic twins. The mode of genetic transmission is unclear. Although some studies have supported the presence of autosomal dominant major loci (3, 4), it has also been argued that bipolar disorder is oligogenic with multiple loci of modest effect.

Although initial attempts at linkage studies met with inconsistent replication (5–8), more recently, the accumulation of multiple studies of larger family sets has led to the reproducible identification of several genetic loci. These include 4p, 12q, 13q, 18, 21q, and Xq among others (9–15). We have previously reported on studies of our set of 20 North American pedigrees, which indicated suggestive evidence of linkage to 5p15 and 22q11 (16, 17). We now report the results of a genome survey using 443 microsatellite markers. These data provide strong evidence for a locus on 22q and support previously reported loci on several other chromosomes. Overlap for some of these loci with those reported for schizophrenia also provides additional data arguing for possible common susceptibility loci for these two disorders.

Methods

Subjects. Families were ascertained from the general North American population through both systematic survey of clinical facilities and through advertising and patient support groups from two sites: San Diego and Vancouver. Families were ascertained through a proband with bipolar I or bipolar II disorder and included if at least two additional members were affected under our broad diagnostic model. Power analyses using SIMLINK (18) were conducted to select both families and members for maximum power. The family sample included 20 families and 164 subjects. Thirty-three subjects had a diagnosis of bipolar I, 15 had bipolar II, and 28 had recurrent major depression. There were averages of 8.2 subjects and 3.8 affected members per family.

Family members were interviewed directly using the Structured Clinical Interview for DSM-3-R (SCID) (19) except for one family, the earliest in the study, which was interviewed using the Schedule for Affective Disorders and Schizophrenia (SADS-L) (20). All diagnoses were made using DSM-3-R criteria, modified to require a 2-day minimum duration for hypomania. Wherever possible, information was also obtained from other family informants and from medical records. Information from all sources was reviewed by a committee of experienced psychiatric clinicians to determine a consensus, best-estimate diagnosis. All interviewers underwent a rigorous, standardized training program for the SCID. Diagnostic reliability was regularly tested by review of videotaped interviews and was consistently high.

DNA samples were obtained from the Coriell Institute for 57 families with bipolar disorder, which were collected as part of the National Institute of Mental Health (NIMH) Bipolar Disorder Genetics Initiative. These families were used as a replication set and were genotyped for selected markers from our genome survey. These 57 families included 345 subjects with the following diagnoses: 169 with bipolar I disorder, 36 with bipolar II disorder, and 45 with recurrent major depression. Ascertainment and diagnostic methods for these families have been described elsewhere (21).

Genotyping. After obtaining informed consent, blood was drawn on all subjects for the establishment of lymphoblastoid cell lines. DNA was prepared from cultured cells by phenol/chloroform extraction.

Markers were primarily tri- and tetranucleotide repeats from the Weber 6.0 screening set (22). These were supplemented with

GENETICS

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: lod, logarithm of odds; cM, centimorgans.

†To whom reprint requests should be addressed. E-mail: jkelsoe@ucsd.edu.

¶Present address: Biofrontera Pharmaceuticals, Leverkusen, Germany.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Article published online before print: *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.011358498. Article and publication date are at www.pnas.org/cgi/doi/10.1073/pnas.011358498

dinucleotide repeat markers from the Genethon map (23). A total of 443 markers was examined. The average intermarker interval was 8 cM; the maximum interval between markers was 33 cM. Map positions were derived primarily from the Marshfield integrated map (24), except for the map used for multipoint analyses on chromosome 22, which was calculated from our genotypic data as described below.

In total, 363 of the microsatellite markers were genotyped at the Novartis laboratory as follows. The PCR was performed in a volume of 20 μl containing 100 ng of genomic DNA, 0.3 μM of each primer (the forward primer had been labeled at the 5′ end with fluorescein), dNTP at a concentration of 130 μM each, 50 mM KCl, 1.5 mM $MgCl_2$, 10 mM Tris-HCl (pH 9.0), and 1 U *Taq* DNA polymerase (Amersham Pharmacia). The PCR amplification was carried out in a Biometra Thermo Cycler with the following cycles: 1×, 5 min at 95°C; 40×, 40 s at 95°C, 40 s at 55°C, and 40 s at 72°C; 1 × 7 min at 72°C. After completion of the amplification, 20 μl of loading buffer (99% of formamide/1 mM EDTA/0.05% bromphenol blue/0.05% xylene cyanol were added to the PCR reaction. The samples were heated up for 3 min at 95°C and then chilled on ice; then, 5 μl of each sample were loaded on a 5% to 8% Long Ranger polyacrylamide/7 M urea gel. The percentage of the gel was dependent on the size of the PCR product. The separated, fluorescently labeled PCR products were visualized by using the Molecular Dynamics FluorImager and ImageQuant software. The instrument scan parameters were adjusted to optimize signal intensity (900 volts, 200-μm pixel size, normal sensitivity).

The remainder of the markers were genotyped at the University of California, San Diego, laboratory as follows: 50 ng of DNA was amplified in either a 5-μl reaction using an ABI 877 catalyst robotic PCR workstation or a 20-μl reaction using an MJ Research PTC-200 thermal cycler. Reactions contained 1.5 mM $MgCl_2$, 50 mM KCl, 200–500 nM of each primer, 10 mM Tris-HCl (pH 8.3), 0.001% gelatin, and 1 unit of either AmpliTaq or AmpliTaq Gold (Perkin–Elmer). The forward primer was labeled with one of three fluors. After an initial 10 min at 95°C (AmpliTaq Gold only), the PCR reaction was conducted using the touchdown protocol (25) in which the annealing temperature was decreased by 1° every two cycles from 65° to 55°, followed by a final 10 cycles at 55°. PCR products were separated by electrophoresis and detected using an ABI 377 and Genescan and Genotyper software. Multiple markers labeled with different fluors and in different molecular weight ranges were pooled along with a molecular weight standard for multiplex detection of between 6 and 12 markers per lane.

All genotypes from each lab were read in a machine-assisted fashion independently by two readers. Standard samples were used for checking consistency between gel molecular weights, and all data were screened for common artifacts using custom data cleaning software. Genotypic data from each laboratory was transferred to University of California, Irvine, for linkage analysis.

Statistical Analysis. A power analysis of this sample was first conducted to aid in the interpretation of results. For this simulation analysis, assumptions were made of an autosomal dominant genetically homogeneous trait and a marker with 4 alleles of equal prevalence, which was 5 cM from the disease gene. Under the narrow disease model, an average logarithm of odds (lod) score of 3.55 was obtained, and there was 65% power to obtain a lod score >3.0. Under the broad model, the average lod score was 5.06, and there was 85% power to detect a lod score >3.0.

Data were recoded to ensure straightforward notation for the multiple alleles in such a way as to allow direct comparison across the two laboratories. Allele frequencies for the markers were estimated from the families using MENDEL (26). These estimates were then compared with published frequencies when the latter were available. Most of the estimates were comparable, and if there were large discrepancies, the pedigrees were reviewed for possible errors. No frequency discrepancies were of sufficient magnitude to warrant multiple runs using different estimates for the same marker(s).

Linkage analyses were carried out using the parametric lod score approach to maximize power over the sample size available. Calculations were performed both for disease vs. marker and then for marker vs. marker combinations in any area that gave some indication of being a potential disease gene location. The marker information was used to confirm the map relations among markers within our data as compared with the published maps (Cooperative Human Linkage Center, etc). One set of data errors for a single chromosome was detected using the map information, and markers were rerun and corrected.

LIPED (27) was used to calculate the two-point lod scores, as that program readily permits calculating the scores using different values for male and female recombination. We modeled the disease in three ways: (i) autosomal dominant with 0.85 penetrance (AD85), (ii), autosomal dominant with 0.50 penetrance (AD50), and (iii) autosomal recessive with 0.50 penetrance (AR50). Markers on the X chromosome were analyzed using a dominant model with 0.85 penetrance. Analyses were conducted under an assumption of genetic homogeneity. Preliminary analyses had indicated that similar results were obtained under heterogeneity as under homogeneity and that the sample did not have adequate power to detect heterogeneity. In addition, we modeled the affected status for bipolar disorder because it is not yet certain which forms may represent the same genetic disease. We defined affected for the narrow diagnostic model (BP only) as bipolar I plus bipolar II plus schizoaffective, bipolar type. The broad diagnostic model (BP + RD) added the category of recurrent major depression to affected status. Subjects with other psychiatric diagnoses were considered of unknown affection status. The three genetic and two diagnostic models resulted in six models that were considered. Penetrance and disease allele frequency was adjusted for each model so as to yield an approximately 5% phenocopy rate and disease prevalences of 1% for the BP only model and 2% for BP + RD. An age-of-onset curve was included with minimal risk below age 15, maximizing at the defined penetrance at age 40.

The lod scores were reviewed by chromosome region to detect inconsistencies. For example, one marker showed some evidence of linkage, but the markers on either side were strikingly negative. Careful review of the pedigrees indicated that the "linked" marker was generated by data from very few informative individuals who by chance had no recombinant offspring. The markers on either side had data from many more informative individuals, and numerous recombinants could be identified. This area was discounted as a potential disease gene region. After the inconsistency review, regions of interest were defined as those with an lod score greater than or equal to 1.0. To increase the density of the map in the regions of interest, additional markers were genotyped.

Three regions were investigated in depth with multipoint analyses using the FASTLINK implementation of the LINKAGE package (28, 29). Two of these regions were on chromosome 22, and one was on chromosome 5. We chose to use only those families that showed potential linkage in each region for the multipoint analyses. Only in one instance was there improvement in the lod score using the multipoint approach.

Results

The results of the genome survey are illustrated in Fig. 1. Twelve markers yielded lod scores greater than 2.0 across the genome. The region with the highest lod scores is on chromosome 22, which is further illustrated in Fig. 2. Thirteen markers spanning

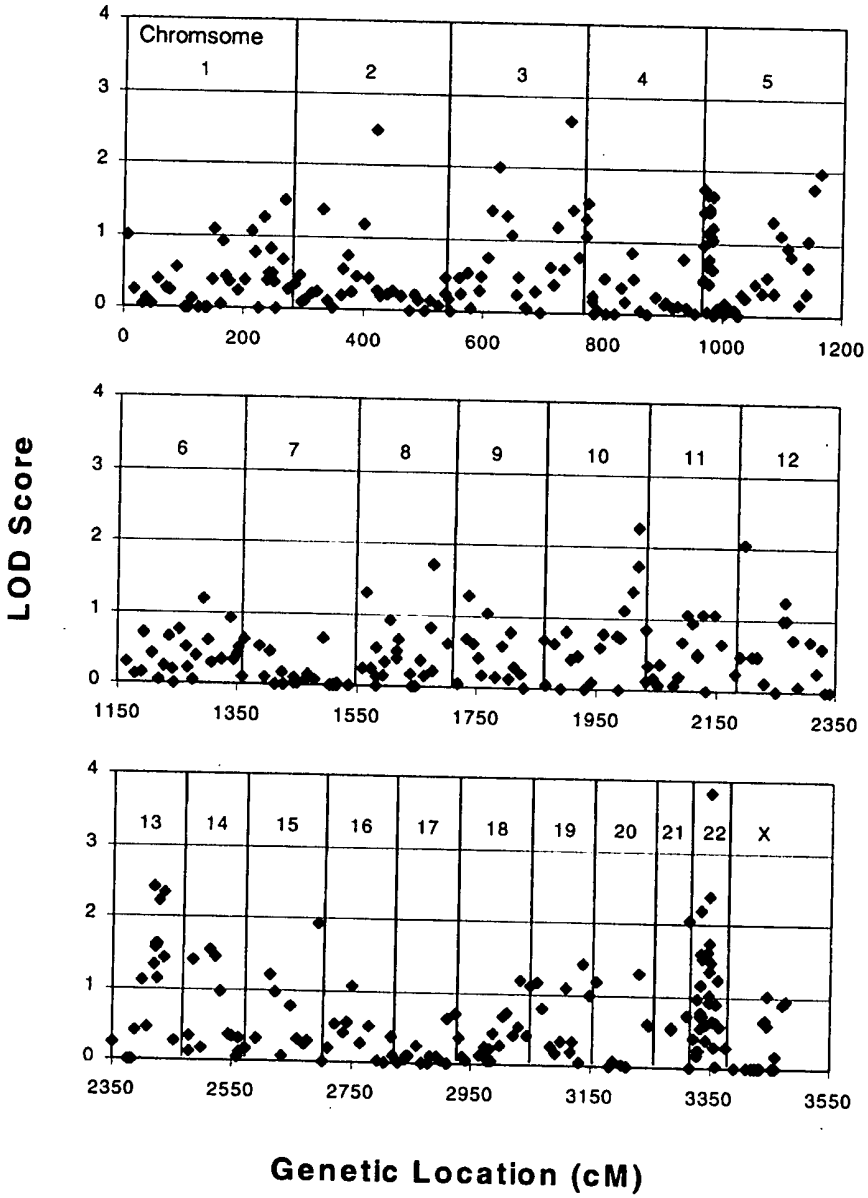
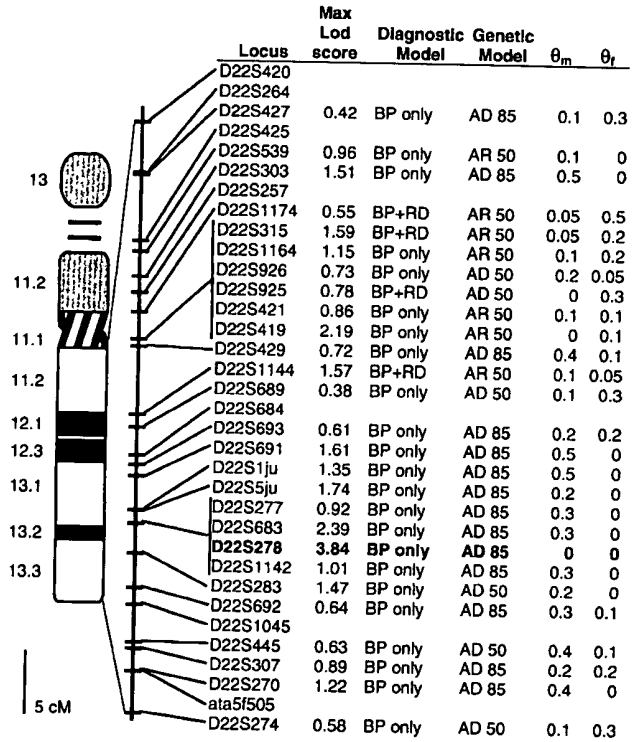


Fig. 1. Four hundred forty-three microsatellite markers were examined spanning the genome. Each point represents the maximum lod score obtained for a microsatellite marker under all models examined. The genetic map was derived primarily from that of the Marshfield laboratory. Genetic location is displayed with chromosomes end to end across the genome. The maximum lod score is displayed as zero if no positive lod score was obtained under any model.

32 cM yielded lod scores greater than 1.0. A maximum lod score of 3.84 was obtained at D22S278 on 22q12. Six other markers within 5 cM also yielded lod scores greater than 1.0, including an lod score of 2.39 at D22S683, which is in close proximity to D22S278. These results were all obtained under the narrow diagnostic model and the high penetrance dominant genetic model. A multipoint analysis was conducted using the eight families with the highest two-point lod scores to further localize the linkage peak. This analysis was consistent with the two-point analysis and yielded a maximum lod score of 3.1 at D22S278. Outside of this immediate region, the next highest lod score was 2.19 at D22S419, 15 cM proximal to D22S278. The length of this overall region of positive lod scores is consistent with other linkage findings in complex disorders (30). Although these results suggest a separate peak near D22S419, it is premature to estimate whether there might be multiple loci involved.

In an attempt to examine these results in an independent sample, we examined 57 families from the NIMH Bipolar Disorder Genetics Initiative sample (21). This is a subset of the larger NIMH family set, which was selected based on size and informativeness of pedigrees. As the results of the NIMH genome scan have already been reported, these results were not intended to provide a new sample for replication but simply to analyze the existing and available NIMH sample using the same markers and analytical methods as for our 20 pedigrees. The results, as illustrated in Table 1, were consistent with those from our genome scan. Eleven of the 16 markers examined on chromosome 22 yielded lod scores greater than 1.0. These positive lod scores spanned the same region implicated in the University of California, San Diego/University of British Columbia (UCSD/UBC) sample. The maximum lod score in the NIMH sample was 2.72 obtained at D22S419. Six other markers



Locus	Max Lod score	Diagnostic Model	Genetic Model	θ_m	θ_f
D22S420					
D22S264					
D22S427	0.42	BP only	AD 85	0.1	0.3
D22S425					
D22S539	0.96	BP only	AR 50	0.1	0
D22S303	1.51	BP only	AD 85	0.5	0
D22S257					
D22S1174	0.55	BP+RD	AR 50	0.05	0.5
D22S315	1.59	BP+RD	AR 50	0.05	0.2
D22S1164	1.15	BP only	AR 50	0.1	0.2
D22S926	0.73	BP only	AD 50	0.2	0.05
D22S925	0.78	BP+RD	AD 50	0	0.3
D22S421	0.86	BP only	AR 50	0.1	0.1
D22S419	2.19	BP only	AR 50	0	0.1
D22S429	0.72	BP only	AD 85	0.4	0.1
D22S1144	1.57	BP+RD	AR 50	0.1	0.05
D22S689	0.38	BP only	AD 50	0.1	0.3
D22S684					
D22S693	0.61	BP only	AD 85	0.2	0.2
D22S691	1.61	BP only	AD 85	0.5	0
D22S1ju	1.35	BP only	AD 85	0.5	0
D22S5ju	1.74	BP only	AD 85	0.2	0
D22S277	0.92	BP only	AD 85	0.3	0
D22S683	2.39	BP only	AD 85	0.3	0
D22S278	**3.84**	**BP only**	**AD 85**	**0**	**0**
D22S1142	1.01	BP only	AD 85	0.3	0
D22S283	1.47	BP only	AD 50	0.2	0
D22S692	0.64	BP only	AD 85	0.3	0.1
D22S1045					
D22S445	0.63	BP only	AD 50	0.4	0.1
D22S307	0.89	BP only	AD 85	0.2	0.2
D22S270	1.22	BP only	AD 85	0.4	0
ata5f505					
D22S274	0.58	BP only	AD 50	0.1	0.3

Fig. 2. Maximum lod score represents the highest two-point lod score obtained for each marker for all models examined. For each marker, the diagnostic and genetic models are indicated, as is the sex-specific recombination fraction that yielded the maximum lod score. No lod score is indicated if no positive lod score was obtained under any model.

within 2 cM also yielded lod scores greater than 1.0. D22S278, which gave the strongest results in the UCSD/UBC sample, yielded a maximum lod score of 1.58. Three other nearby markers also obtained lod scores greater than 1.0. Therefore, for each of the two putative linkage peaks on 22q, positive scores were obtained in both samples. However, the region near D22S419 was strongest in the NIMH sample, whereas the region near D22S278 was strongest in the UCSD/UBC sample.

Table 2 details several other genomic regions that yielded suggestive evidence of linkage in the genome scan of the UCSD/UBC family set. Each of these regions includes at least one marker with an lod score greater than 2.0, which is flanked by other markers with lod scores greater than 1.0. Two separate regions on chromosome 3 were implicated. Each region is approximately 30 cM long, and the regions are separated by 60 cM. D10S1223 yielded an lod score of 2.27 and is flanked by several markers with lod scores greater than 1.0 over an interval of 20 cM. Another region with particularly suggestive data is 13q. Three markers, D13S154, D13S225, and D13S796, generated lod scores greater than 2.0. Six other makers in this 20-cM interval also yielded lod scores greater than 1.0. On 21q, the marker, PFKL, yielded a lod score of 2.04.

5p15 is of particular interest because it includes an important candidate gene, the dopamine transporter. Although the combined family sample yielded only modest evidence of linkage, one large family (family 16) generated suggestive evidence of linkage. As shown in Table 3, a maximum lod score of 2.77 was obtained for this family at D5S417. These results are consistent with those we have previously reported in the same family set with a less extensive set of markers (16).

Several other regions deserve mention. Lod scores above 2.0 were detected on chromosomes 2q and 12p. However, the absence of any supporting evidence from immediately flanking markers caused us to dismiss the importance of these results. Several other regions yielded modestly positive lod scores in the range of 1.0 to 2.0 and are interesting because of other reports in these regions. These include 4p, 5q, and 16p.

Discussion

We examined 443 microsatellite markers in a set of 20 North American families with bipolar disorder. The genome-wide maximum was a lod score of 3.8 at D22S278 on 22q13 under the narrow diagnostic definition and an autosomal dominant model. Other regions with suggestive evidence for linkage include 3p21, 3q27, 5p15, 10q, 13q31-q34, and 21q22.

We first reported suggestive evidence of linkage to 22q in 1997 in a subset of this sample (31). This earlier study examined 13 of the 20 families reported here and obtained a maximum lod score of 2.51 at D22S303. This report extends our results by examining an expanded number of markers in a larger family set. In the current sample, the maximum evidence of linkage is found approximately 20 cM distal to the original result; however, D22S303 is near the possible second peak of linkage reported here at D22S419. Our results are consistent with several other reports examining this region. The NIMH Bipolar Disorder Genetics Initiative consortium reported an lod score of 2.5 at

Table 1. Two-point lod scores for bipolar disorder in 57 NIMH families

Locus	Position (cM)	Max LOD score*	Diagnostic model	Genetic model	θ_m	θ_f
D22S303	16.4	0.83	BP only	AR 50	0.5	0.2
D22S1174	19.3	0.86	BP only	AD 50	0.5	0.2
D22S315	21.5	1.14	BP only	AR 50	0.5	0.2
D22S925	21.5	2.25	BP only	AR 50	0.5	0.1
D22S421	21.5	1.12	BP only	AR 50	0.5	0.2
D22S419	21.5	2.72	BP only	AR 50	0.5	0.1
D22S533	22.0	1.17	BP only	AR 50	0.3	0.1
D22S1144	27.5	1.85	BP only	AR 50	0.5	0.05
D22S689	28.6	2.17	BP only	AR 50	0.5	0.05
D22S691	32.4	0.82	BP only	AR 50	0.5	0.1
D22S1ju	35.2	1.46	BP only	AR 50	0.4	0.1
D22S683	36.2	1.34	BP only	AR 50	0.4	0.1
D22S278	36.2	1.58	BP + RD	AD 50	0.5	0.2
D22S283	38.6	1.00	BP only	AR 50	0.4	0.2
D22S692	41.4	0.28	BP only	AR 50	0.5	0.3
D22S1045	42.8	0.63	BP only	AD 85	0.5	0.3

*The maximum lod score obtained under any model tested.

Table 2. Two-point lod scores from other regions with suggestive evidence of linkage

Chrom	Locus	Position (cM)	Max LOD Score*	Diagnostic model	Genetic model	θ_m	θ_f
3	D3S2409	70.61	0.74	BP only	AR50	0.001	0.5
3	D3S1766	78.64	1.40	BP only	AD85	0.001	0.3
3	D3S4542	89.91	2.01	BP only	AD85	0.2	0.001
3	D3S2406	102.64	1.32	BP only	AD85	0.1	0.2
3	GATA128C02	112.42	1.07	BP only	AD50	0.05	0.2
3	D3S2459	119.09	0.25	BP only	AD85	0.2	0.3
3	D3S3053	181.87	0.38	BP only	AD85	0.001	0.4
3	D3S2427	188.29	1.18	BP only	AD85	0.4	0.001
3	D3S1602	201.14	0.61	BP only	AD85	0.3	0.1
3	D3S2398	209.41	2.66	BP only	AD85	0.2	0.001
3	D3S2418	215.84	1.41	BP only	AR50	0.5	0.001
3	D3S1311	224.88	0.78	BP only	AD50	0.05	0.4
10	D10S1237	134.7	1.11	BP + RD	AD85	0.05	0.5
10	D10S587	147.57	1.37	BP only	AD50	0	0.3
10	D10S1223	156.27	2.27	BP only	AR50	0.001	0.5
10	D10S217	157.89	1.74	BP only	AR50	0	0.5
13	D13S154	75.19	2.4	BP + RD	AR50	0.01	0.4
13	D13S793	76.26	1.32	BP only	AR50	0.05	0.5
13	D13S770	79.49	1.62	BP + RD	AD50	0.01	0.5
13	D13S128	79.49	1.56	BP only	AR50	0.05	0.5
13	D13S1240	81.64	1.61	BP only	AR50	0.001	0.5
13	D13S779	82.93	1.14	BP only	AD50	0.001	0.05
13	D13S225	83.57	2.22	BP only	AR50	0.001	0.5
13	D13S173	93.52	1.41	BP + RD	AR50	0.05	0.5
13	D13S796	93.52	2.34	BP + RD	AD85	0.01	0.3
21	PFKL	65.6	2.04	BP + RD	AD50	0.4	0.001

*The maximum lod score obtained under any model tested.

D22S533 using a multipoint sib pair analysis (32). This is approximately 1 cM from our secondary peak at D22S419. The 57 NIMH families we examined are a subset of the NIMH Genetics Initiative pedigrees, and our results are consistent with theirs. These results on chromosome 22 are also consistent with a separate set of 21 families collected by the NIMH intramural program. Detera-Wadleigh *et al.* (11) reported a maximum lod score of 2.5 using a multipoint sib pair analysis in the interval between D22S689 and D22S685. Linkage maps position this region approximately halfway between the peaks at D22S419 and D22S278 found in our study. However, more recent genomic sequence and physical mapping data indicate that this region is more telomeric, approximately 2 Mb centromeric to D22S278 (33).

This 22q region has also been implicated in several studies of schizophrenia. Pulver *et al.* (34) first reported evidence of linkage of schizophrenia to 22q at the IL2RB locus near D22S278. Subsequently, this region has been implicated in several other linkage studies and the marker D22S278 in several linkage disequilibrium studies (35–38). Although it is this region near D22S278 that has been primarily reported in studies of schizophrenia, Myles-Worsley *et al.* have recently reported evidence of linkage of schizophrenia to a marker near our possible secondary peak at D22S419 (39). They examined a composite inhibitory endophenotype in families with schizophrenia and found a genome-wide maximum lod score of 3.5 at D22S315, which is very close to D22S419. Together, these data provide a considerable amount of evidence for loci for both bipolar disorder and schizophrenia on 22q. Furthermore, the data for both disorders suggest the possibility of more than one linkage peak on this chromosome.

Other regions of suggestive evidence for linkage in our study are consistent with previous reports and provide supportive evidence for these loci. 13q has been implicated in a recent study

Table 3. Two-point lod scores at 5p15 in family 16

Locus	Position (cM)	Max lod score*	Diagnostic model	Genetic model	θ_m	θ_f
D5S2005	1.72	2.08	BP only	AD85	0.001	0.001
D5S678	1.72	2.14	BP only	AD85	0.001	0.001
D5S1981	1.72	1.89	BP only	AD85	0.001	0.001
D5S1970	5.43	0			0.5	0.5
D5S417	6.67	2.77	BP + RD	AD85	0.05	0.1
D5S675	9.41	1.57	BP + RD	AD85	0.05	0.5
D5S405	9.41	1.64	BP only	AD85	0.001	0.001
D5S2088	9.41	0.70	BP only	AR85	0.001	0.001
D5S1492	9.41	2.02	BP + RD	AD85	0.001	0.2
D5S406	11.85	2.01	BP + RD	AD85	0.001	0.001
D5S2054	14.3	1.84	BP only	AD85	0.001	0.001
D5S464	14.3	1.88	BP only	AD85	0.001	0.001
D5S2505	14.3	1.51	BP only	AD85	0.001	0.001
D5S635	14.91	1.45	BP only	AD85	0.001	0.001

*The maximum lod score obtained under any model tested.

of bipolar disorder by Detera-Wadleigh *et al.* (11) who reported a lod score of 3.5 in this region. Our data represent a significant replication of these results (40). Like 22q, this region has also been implicated in studies of schizophrenia (41). Evidence for linkage of bipolar disorder to the PFKL locus on 21q was reported in 1994 by Straub *et al.* (14). Subsequently, this locus has been supported by several other studies, and our data provide further support for this result. Markers on 10q were implicated by the NIMH Genetics Initiative consortium who found evidence for linkage at D10S188, about 20 cM proximal to the positive region in our data. Linkage has also been reported in a German sample of bipolar families at D10S217 (S. Cichon, personal communication). Two other groups have also reported evidence of linkage to schizophrenia in this region (42). Few studies of bipolar disorder have implicated chromosome 3; however, Edenberg *et al.* have reported increased allele sharing at D3S3038 in studies of the NIMH Genetics Initiative for Bipolar Disorder sample (32).

Our results must be qualified based on the number of different models examined. We chose a parametric approach because of its greater power, and we chose to use a limited number of both diagnostic and genetic models to best cover the most likely modes of transmission. The complexity and non-independence of the models used make it difficult to estimate the exact significance of these results. Rather, the strength of our results must be interpreted relative to other regions in our genome survey and in the context of supporting data in our examination of the NIMH sample and other independent data sets.

Three of the regions implicated in this study (22q, 13q, and 10q) have also been reported in studies of schizophrenia. Such a correspondence has also been reported for 18p and 10p (43). Although such an overlap could occur by chance or reflect two nearby but separate susceptibility genes, it also raises the intriguing possibility that many susceptibility genes are common for the two disorders. Although family studies have generally found these two disorders to "breed true," a small degree of familial overlap has frequently been observed. Recent family study data have further supported some degree of overlap (44). Only the identification of specific genes will resolve this question. However, these data raise the hypothesis that some genes contribute to susceptibility in a nonspecific way or that different mutations in the same gene may predispose to different illnesses.

In summary, our results support the presence of a susceptibility locus for bipolar disorder on chromosome 22. They also provide support for regions previously reported on 5p, 10q, 13q, and 21q and suggestive evidence for novel loci on 3p and 3q. It is intriguing that three of these regions have also been implicated in studies of schizophrenia. These molecular data raise the possibility that common susceptibility genes may be involved and that the relationship between these two disorders may be more complex than previously thought.

We thank the family members who participated in this study, without whom it would not be possible. This work was supported by Novartis Pharma AG and by grants (to J.R.K.) from the Department of Veterans Affairs and the National Institute of Mental Health Grants MH47612 and MH59567. Support was also provided by the University of California, San Diego, Mental Health Clinical Research Center (MH30914) and General Clinical Research Center (M01 RR00827). Data and biomaterials were collected in four projects that participated in the National Institute of Mental Health Bipolar Disorder Genetics Initiative. From 1991 to 1998, the Principal Investigators and Co-Investigators were Indiana University, Indianapolis, IN (UO1 MH46282, John Nurnberger, M.D., Ph.D., Marvin Miller, M.D., and Elizabeth Bowman, M.D.); Washington University, St. Louis, MO (UO1 MH46280, Theodore Reich, M.D., Allison Goate, Ph.D., and John Rice, Ph.D.); Johns Hopkins University, Baltimore, MD (UO1 MH46274, J. Raymond DePaulo, Jr., M.D., Sylvia Simpson, M.D., MPH, and Cohn Stine, Ph.D.); and the National Institute of Mental Health Intramural Research Program, Clinical Neurogenetics Branch, Bethesda, MD (Elliot Gershon, M.D., Diane Kazuba, B.A., and Elizabeth Maxwell, M.S.W.).

1. Goodwin, F. K. & Jameson, K. R. (1990) *Manic-Depressive Illness* (Oxford Univ. Press, New York).
2. Tsuang, M. T. & Faraone, S. V. (1990) *The Genetics of Mood Disorders* (Johns Hopkins Univ. Press, Baltimore).
3. Spence, M. A., Flodman, P. L., Sadovnick, A. D., Bailey-Wilson, J. E., Ameli, H. & Remick, R. A. (1995) *Am. J. Med. Genet.* **60**, 370–376.
4. Rice, J., Reich, T., Andreasen, N. C., Endicott, J., van Eerdewegh, M., Fishman, R., Hirschfeld, R. M. & Klerman, G. L. (1987) *Arch. Gen. Psychiatry* **44**, 441–447.
5. Egeland, J. A., Gerhard, D. S., Pauls, D. L., Sussex, J. N., Kidd, K. K., Allen, C. R., Hostetter, A. M. & Housman, D. E. (1987) *Nature (London)* **325**, 783–787.
6. Kelsoe, J. R., Ginns, E. I., Egeland, J. A., Gerhard, D. S., Goldstein, A. M., Bale, S. J., Pauls, D. L., Long, R. T., Kidd, K. K., Conte, G., *et al.* (1989) *Nature (London)* **342**, 238–243.
7. Baron, M., Risch, N., Hamburger, R., Mandel, B., Kushner, S., Newman, M., Drumer, D. & Belmaker, R. H. (1987) *Nature (London)* **326**, 289–292.
8. Baron, M. (1991) *Soc. Biol.* **38**, 179–188.
9. Blackwood, D. H., He, L., Morris, S. W., McLean, A., Whitton, C., Thomson, M., Walker, M. T., Woodburn, K., Sharp, C. M., Wright, A. F., *et al.* (1996) *Nat. Genet.* **12**, 427–430.
10. Dawson, E., Parfitt, E., Roberts, Q., Daniels, J., Lim, L., Sham, P., Nothen, M., Propping, P., Lanczik, M., Maier, W., *et al.* (1995) *Am. J. Med. Genet.* **60**, 94–102.
11. Detera-Wadleigh, S. D., Badner, J. A., Berrettini, W. H., Yoshikawa, T., Goldin, L. R., Turner, G., Rollins, D. Y., Moses, T., Sanders, A. R., Karkera, J. D., *et al.* (1999) *Proc. Natl. Acad. Sci. USA* **96**, 5604–5609.
12. Berrettini, W. H., Ferraro, T. N., Goldin, L. R., Weeks, D. E., Detera-Wadleigh, S., Nurnberger, J. I. J. & Gershon, E. S. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 5918–5921.
13. Freimer, N. B., Reus, V. I., Escamilla, M. A., McInnes, L. A., Spesny, M., Leon, P., Service, S. K., Smith, L. B., Silva, S., Rojas, E., *et al.* (1996) *Nat. Genet.* **12**, 436–441.
14. Straub, R. E., Lehner, T., Luo, Y., Loth, J. E., Shao, W., Sharpe, L., Alexander, J. R., Das, K., Simon, R., Fieve, R. R., *et al.* (1994) *Nat. Genet.* **8**, 291–296.
15. Pekkarinen, P., Terwilliger, J., Bredbacka, P. E., Lonnqvist, J. & Peltonen, L. (1995) *Genome Res.* **5**, 105–115.
16. Kelsoe, J. R., Sadovnick, A. D., Kristbjarnarson, H., Bergesch, P., Mroczkowski-Parker, Z., Drennan, M. D., Rapaport, M. H., Flodman, P., Spence, M. A. & Remick, R. A. (1996) *Am. J. Med. Genet.* **67**, 533–540.
17. Lachman, H. M., Kelsoe, J. R., Remick, R. A., Sadovnick, A. D., Rapaport, M. H., Lin, M., Pazur, B. A., Roe, A. A., Saito, T. & Papolos, D. F. (1996) *Am. J. Med. Genet.* **74**, 121–128.
18. Ploughman, L. M. & Boehnke, M. (1989) *Am. J. Hum. Genet.* **44**, 543–551.
19. Spitzer, R. L., Williams, J. B. W. & Gibbon, M. (1987) *Structured Clinical Interview for DSM-3-R(SCID)* (New York State Psychiatric Institute Biometrics Research, New York).
20. Endicott, J. & Spitzer, R. (1978) *Arch. Gen. Psychiatry* **35**, 837–844.
21. Nurnberger, J. I. J., DePaulo, J. R., Gershon, E. S., Reich, T., Blehar, M. C., Edenberg, H. J., Foroud, T., Miller, M., Bowman, E., Mayeda, A., *et al.* (1997) *Am. J. Med. Genet.* **74**, 227–237.
22. Murray, J. C., Buetow, K. H., Weber, J. L., Ludwigsen, S., Scherpbier-Heddema, T., Manion, F., Quillen, J., Sheffield, V. C., Sunden, S. & Duyk, G. M. (1994) *Science* **265**, 2049–2054.
23. Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., *et al.* (1996) *Nature (London)* **380**, 152–154.
24. Broman, K. W., Murray, J. C., Sheffield, V. C., White, R. L. & Weber, J. L. (1998) *Am. J. Hum. Genet.* **63**, 861–869.
25. Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K. & Mattick, J. S. (1991) *Nucleic Acids Res.* **19**, 4008.
26. Lange, K., Weeks, D. & Boehnke, M. (1988) *Genet. Epidemiol.* **5**, 471–472.
27. Ott, J. (1974) *Am. J. Hum. Genet.* **26**, 588–597.
28. Cottingham, R. W., Jr., Idury, R. M. & Schaffer, A. A. (1993) *Am. J. Hum. Genet.* **53**, 252–263.
29. Lathrop, G. M., Lalouel, J. M., Julier, C. & Ott, J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3443–3446.
30. Terwilliger, J. D., Shannon, W. D., Lathrop, G. M., Nolan, J. P., Goldin, L. R., Chase, G. A. & Weeks, D. E. (1997) *Am. J. Hum. Genet.* **61**, 430–438.
31. Lachman, H. M., Kelsoe, J. R., Remick, R. A., Sadovnick, A. D., Rapaport, M. H., Lin, M., Pazur, B. A., Roe, A. M., Saito, T. & Papolos, D. F. (1997) *Am. J. Med. Genet.* **74**, 121–128.
32. Edenberg, H. J., Foroud, T., Conneally, P. M., Sorbel, J. J., Carr, K., Crose, C., Willig, C., Zhao, J., Miller, M., Bowman, E., *et al.* (1997) *Am. J. Med. Genet.* **74**, 238–246.
33. Dunham, I., Shimizu, N., Roe, B. A., Chissoe, S., Hunt, A. R., Collins, J. E., Bruskiewich, R., Beare, D. M., Clamp, M., Smink, L. J., *et al.* (1999) *Nature (London)* **402**, 489–495.
34. Pulver, A. E., Karayiorgou, M., Wolyniec, P. S., Lasseter, V. K., Kasch, L., Nestadt, G., Antonarakis, S., Housman, D., Kazazian, H. H. & Meyers, D. (1994) *Am. J. Med. Genet.* **54**, 36–43.
35. Coon, H., Jensen, S., Holik, J., Hoff, M., Myles-Worsley, M., Reimherr, F., Wender, P., Waldo, M., Freedman, R. & Leppert, M. (1994) *Am. J. Med. Genet.* **54**, 59–71.
36. Vallada, H., Curtis, D., Sham, P. C., Murray, R. M., McGuffin, P., Nanko, S., Gill, M., Owen, M. & Collier, D. A. (1995) *Psychiatr. Genet.* **5**, 127–130.
37. Moises, H. W., Yang, L., Li, T., Havsteen, B., Fimmers, R., Baur, M. P., Liu, X. & Gottesman, I. I. (1995) *Am. J. Med. Genet.* **60**, 465–467.
38. Schizophrenia Collaborative Linkage Group for Chromosome 22. (1998) *Schizophr. Res.* **32**, 115–121.
39. Myles-Worsley, M., Coon, H., McDowell, J. E., Brenner, C., Hoff, M., Lind, B., Rosenthal, J., Bennett, P., Freedman, R., Clementz, B. A. & Byerley, W. (1999) *Am. J. Med. Genet.* **88**, 544–550.
40. Lander, E. & Kruglyak, L. (1995) *Nat. Genet.* **11**, 241–247.
41. Blouin, J. L., Dombroski, B. A., Nath, S. K., Lasseter, V. K., Wolyniec, P. S., Nestadt, G., Thornquist, M., Ullrich, G., McGrath, J., Kasch, L., *et al.* (1998) *Nat. Genet.* **20**, 70–73.
42. Wildenauer, D. B. & Schwab, S. G. (1999) *Am. J. Med. Genet.* **88**, 239–243.
43. Berrettini, W. H. (2000) *Biol. Psychiatry* **47**, 245–251.
44. Erlenmeyer-Kimling, L., Adamo, U. H., Rock, D., Roberts, S. A., Bassett, A. S., Squires-Wheeler, E., Cornblatt, B. A., Endicott, J., Pape, S. & Gottesman, I. I. (1997) *Arch. Gen. Psychiatry* **54**, 1096–1102.

Exhibit P

Nerve Growth Factor–specific Regulation of Protein Methylation during Neuronal Differentiation of PC12 Cells

Thomas R. Cimato,* Murray J. Ettinger,* Xianbo Zhou,* and John M. Aletta‡

*Department of Biochemistry, ‡Department of Pharmacology and Toxicology, University at Buffalo School of Medicine and Biomedical Sciences, State University of New York, Buffalo, New York 14214

Abstract. Protein methylation is a posttranslational modification that can potentially regulate signal transduction pathways in a similar manner as protein phosphorylation. The role of protein methylation in NGF signaling was examined by metabolic labeling of PC12 cell proteins with L-[methyl-^{3}H]methionine and by in vitro labeling of cell proteins with L-[methyl-^{3}H]S-adenosylmethionine. Effects of NGF were detected within 15 min. Methyl-labeled proteins were resolved by one and two dimensional SDS-PAGE. NGF affected the methylation of several 68–60-kD proteins (pI 5.8–6.4) and 50-kD proteins (isoelectric point pH 6.7–6.8 and 5.8–6.2). Several NGF-induced changes in methylation increased over several hours and through 4 d. Moreover, methyl labeling of several specific proteins was only detected after NGF treatment, but not in non-treated controls. The effects of NGF on protein methylation were NGF specific since they were not observed with EGF or insulin. A requirement for protein methylation for neurite outgrowth was substantiated with either of two methylation inhibitors: dihydroxycyclopentenyl adenine (DHCA) and homocysteine. DHCA, the more potent of the two, markedly inhibits protein methylation and neurite outgrowth without affecting cell growth, NGF-induced survival, cell flattening, or several protein phosphorylations that are associated with early signaling events. Removal of DHCA leads to rapid protein methylation of several proteins and concurrent neurite outgrowth. The results indicate that NGF regulates the methylation of several specific proteins and that protein methylation is involved in neurite outgrowth from PC12 cells.

PROTEIN methylation is a posttranslational modification that may be used to regulate signal transduction and differentiation pathways by mechanisms that are analogous to regulation by protein phosphorylation (Hrycyna and Clarke, 1993; Rando, 1996). Although a specific role for protein methylation in prokaryote chemotaxis is well established (Shapiro et al., 1995), possible roles for protein methylation in eukaryotic signaling mechanisms have not been extensively explored. Protein carboxyl methylations are reversible, and regulatory roles for carboxyl methylation have been proposed for chemoattractant responses in neutrophils (Philips et al., 1993, 1995), insulin secretion from pancreatic islets (Metz et al., 1993), and photoreceptor signal transduction (Parish et al., 1995). Among signaling proteins that are known to be carboxyl methylated are the Ras and Rho family of small G-proteins (Hrycyna and Clarke, 1993; Rando, 1996), γ subunits of heterotrimeric G-proteins (Philips et al., 1993; Rando, 1996), and the catalytic subunit of protein phosphatase 2A (Lee and Stock, 1993; Favre et al., 1994; Xie and Clarke, 1994). Protein phosphatase 2A is also demethylated by a specific protein carboxyl methylesterase (Lee et al., 1996). Many proteins are known to be N-methylated at arginine, lysine, or histidine residues including cytoskeletal proteins (actin and myosin), nuclear proteins (nucleolin, fibrillarin, histones, heterogeneous nuclear RNPs), the multifunctional calcium binding protein, calmodulin, and FGF-2. The physiological functions of methylation, however, remain largely unknown. Moreover, potential mechanisms for regulating protein methylation within growth factor signaling pathways remain to be explored.

Methylation pathways use S-adenosylmethionine (SAM)[1] as the universal methyl donor for methyltransferase-catalyzed methylation of proteins and other methyl acceptors. The role of methylation in NGF signal transduction was previously studied using high concentrations (millimolar) of inhibitors of methyltransferases that use SAM (Seeley et al., 1984; Kujubu et al., 1993). These attempts to exam-

Please address all correspondence to John M. Aletta, Department of Pharmacology and Toxicology, University at Buffalo School of Medicine and Biomedical Sciences, 3435 Main Street, Buffalo, NY 14214-3000. Tel.: (716) 829-3237. Fax: (716) 829-2801. e-mail: jaletta@ubmed.buffalo.edu

1. *Abbreviations used in this paper*: DHCA, 9-(*trans*-2′, *trans*-3′-dihydroxycyclopent-4′-enyl)-adenine; ERK, extracellular signal-regulated kinase; PAS, protein-A-Sepharose; 2D two dimensional; PI, phosphoinositide; SAHcy, S-adenosylhomocysteine; SAHH, S-adenosylhomocysteine hydrolase; SAM, S-adenosylmethionine; Trk, tyrosine kinase receptor.

The Journal of Cell Biology, Volume 138, Number 5, September 8, 1997 1089–1103
http://www.jcb.org

ine the role of methylation in cell signaling were somewhat compromised by the lack of specific, nontoxic inhibitors of methylation. In addition to the methylated acceptor, S-adenosylhomocysteine (SAHcy) is a product of all methyltransferase reactions. SAHcy is a strong competitive inhibitor of SAM (Hildeshein et al., 1972) and is normally removed by hydrolysis in a reversible reaction (De la Haba and Cantoni, 1959) catalyzed by S-adenosylhomocysteine hydrolase (SAHH). Thus, inhibition of SAHH offers an alternative means to inhibit methyltransferases. This approach was used in the present work to study protein methylation during NGF signaling. The crucial role of SAHH and methylation in development is evident from the nonagouti (a^x) mutation. The lethality of this mutation is due to a deletion of the SAHH gene, and embryonic development is arrested in the preimplantation blastula stage (Miller et al., 1994).

The PC12 clonal cell line was used to examine the role of protein methylation in NGF signal transduction. PC12 cells are derived from a rat pheochromocytoma and serve as a model of neuronal differentiation. When exposed to NGF, PC12 cells assume many of the features of sympathetic neurons including cell cycle arrest, survival in serum-free medium, and elaboration of long neurites (Greene and Tischler, 1976; Tischler and Greene, 1978). Moreover, PC12 cells also respond to other growth factors. For example, in the presence EGF, PC12 cells do not differentiate into neuronlike cells, but do increase their mitotic rate (Huff and Guroff, 1981). Thus, PC12 cells are useful to study specific growth factor signaling mechanisms. The specific signaling events and proteins involved in neurite outgrowth have not been fully elucidated. For example, NGF and EGF both activate several signaling proteins including phospholipase C, phosphoinositide (PI)3-kinase, extracellular signal–related kinase (ERKs) (MAP kinases), and Ras (Kaplan and Stephens, 1994; Marshall, 1995). However, NGF, but not EGF, stimulates neurite outgrowth in PC12 cells. One hypothesis for the specificity of NGF actions involves kinetic differences in the activation of both Ras and ERK (Marshall, 1995). NGF stimulation of PC12 cells results in prolonged activation of both Ras and ERK activity, while EGF causes a transient rise and fall in Ras and ERK activities (Qui and Green, 1992). Specificity may also arise from NGF-specific signaling pathways that are not activated by EGF and other mitogens (Chao, 1992; Kaplan and Stephens, 1994; Peng et al., 1995). The results reported here indicate that signaling pathways involving protein methylation may contribute to NGF specificity.

Previous studies by Seeley et al. (1984) have implied that methylation is necessary for neurite outgrowth in PC12 cells. Kujubu et al. (1993) and Haklai et al. (1993) reported that methylation of small G-proteins is regulated by NGF in PC12 cells. Najbauer and Aswad (1990) have also characterized several methyl-arginine containing proteins in PC12 cells. The effects of NGF on protein methylation and the effect of inhibition of protein methylation on neurite outgrowth are, however, incompletely characterized and have been difficult to interpret due to the lack of specificity of the methylation inhibitors used in previous studies. The present work extends previous findings by examining the methylation of specific cellular proteins and neurite outgrowth before and after inhibition of the protein methylation pathway. 9-(*trans*-2′, *trans*-3′-dihydroxycyclopent-4-enyl)-adenine (DHCA), a specific, mechanism-based inhibitor of SAHH, inhibits protein methylation and greatly decreases neurite outgrowth. The results indicate that NGF, but not EGF or insulin, regulates the methylation of several specific proteins and that protein methylation is required for neurite outgrowth from PC12 cells.

Materials and Methods

Reagents

NGF was purified from male mouse submaxillary glands as described previously (Mobley et al., 1976). Insulin, L-homocysteine thiolactone, anisomycin, leupeptin, PMSF, aprotinin, and monoclonal phosphotyrosine antibody (clone PT-66) were purchased from Sigma Chemical Co. (St. Louis, MO). Homocysteine was prepared by incubating L-homocysteine thiolactone with 40 mM NaOH at 37°C for 30 min. EGF was obtained from Upstate Biotechnology Inc. (Lake Placid, NY). DHCA (kindly provided by R.T. Borchardt, University of Kansas, Lawrence, KS) was prepared as 1 and 100 mM stock solutions in DMSO. Erythro-9-(2-hydroxy-3-nonyl) adenine was kindly provided by D. Porter (Burroughs Wellcome, Research Triangle Park, NC). L-[methyl-^{3}H]methionine (71.4 Ci/mmol), L-[methyl-^{3}H]SAM (70 Ci/mmol), [^{14}C]adenosine (59.8 mCi/mmol), and carrier-free [^{32}P]orthophosphate were purchased from Dupont-NEN (Boston, MA). ^{125}I-donkey anti-rabbit IgG was purchased from Amersham Corp. (Arlington Heights, IL). Protein A–Sepharose (PAS) and ampholines were obtained from Pharmacia Fine Chemicals (Piscataway, NJ). Donor horse serum and fetal bovine serum were from JRH Biosciences (Lenaxa, KS). Cell culture medium, penicillin, and streptomycin were obtained from Gibco Laboratories (Grand Island, NY). Restriction endonuclease MspI was from New England Biolabs Inc. (Beverly, MA), and nuclease P1 from United States Biochemical Corp. (Cleveland, OH).

Cell Culture and Bioassays

Stock cultures of PC12 cells were grown on collagen-coated tissue culture dishes (Falcon Plastics, Cockeysville, MD) in RPMI-1640 medium supplemented with 10% heat-inactivated horse serum and 5% fetal bovine serum at 35°C and 7.5% CO_2, as previously described (Greene et al., 1987). For neurite outgrowth studies, PC12 cells (3–5 × 10^5 cells) were plated onto collagen-coated 35-mm plastic tissue culture dishes and cultured in RPMI-1640 medium containing 1% heat-inactivated horse serum and 50 ng/ml NGF. Neurites >20 μm were counted as neurite-bearing cells. At least 100 cells from each experimental condition were scored from randomly chosen fields.

For the study of proliferation and survival in the presence and absence of DHCA, naive PC12 cells were plated on collagen-coated, 24-well tissue culture dishes at a density of 10^5 cells per well. 1 d after plating, cells were washed three times in RPMI-1640 medium to remove serum and then placed either in complete medium (RPMI plus 10% horse serum and 5% fetal bovine serum), in RPMI plus 50 ng/ml NGF, or in RPMI alone. Where indicated, 1 μM DHCA was included in the culture medium. The number of viable cells was determined by counting intact nuclei using a hemacytometer (Soto and Sonnenschein, 1985) immediately after washing the cultures with RPMI at time intervals of 1, 3, and 5 d later.

For the study of neurite regeneration, PC12 cells were treated with NGF for 7–14 d. NGF was then thoroughly washed away from the cells, and the cells were detached from the culture dish by trituration. Cells were then replated in the presence or absence of NGF, and neurite outgrowth was scored 24 h later.

Metabolic Radiolabeling of Proteins

To assess endogenous protein methylation under various experimental conditions, PC12 cells were cultured in 1% horse serum for at least 16 h before metabolic radiolabeling for 6 h with 100 μCi/ml L-[methyl-^{3}H]methionine in the presence of 10 μM anisomycin at 35°C in a CO_2 incubator. The methionine/cysteine content of RPMI-1640 was reduced by 80% during the labeling period. Growth factors (NGF, EGF, or insulin) and/or 1 μM DHCA were added for the indicated times. At the conclusion of the labeling period, cells were washed three times with 1 ml of PBS (35°C),

harvested in 200 μl SDS-PAGE lysis buffer, and held in a boiling water bath for 5 min. Total radiolabeled protein in each lysate was determined by TCA precipitation and liquid scintillation spectrometry. The averages of the determinations are given under Results as the means ± SEM. Aliquots of each cell lysate containing equal TCA-precipitable counts per minute were subjected to SDS-PAGE as described below. For analysis of protein methylation by two-dimensional (2D) IEF × SDS-PAGE, the cells were harvested in IEF lysis buffer (Aletta and Greene, 1987) without heating the sample.

To assess changes in protein phosphorylation, PC12 cell proteins were metabolically radiolabeled as previously described (Aletta, 1996) in a Hepes-buffered Krebs-Ringer solution containing 0.1% D-glucose. When the effect of DHCA was examined, the drug was added to cultures 30 min before addition of the radioisotope. The labeling was carried out with 50 μCi/ml [^{32}P]orthophosphate at 35°C in room air for 2 h. At the end of the labeling period, cells were washed three times with 1 ml of PBS (35°C), harvested in SDS-PAGE lysis buffer, and held in a boiling water bath for 5 min. Total radiolabeled protein in each lysate was determined by TCA precipitation and liquid scintillation spectometry.

In Vitro Protein Methylation

To measure protein methylation in subcellular fractions, cells were washed three times in PBS (4°C) and harvested by scraping with a rubber policeman in ice-cold homogenization buffer (100 mM Tris, pH 8.0, 1 mM EDTA, 2 μM PMSF, and 10 μg/ml leupeptin). After homogenization at 4°C in a Dounce homogenizer, the homogenates were centrifuged at 4°C for 10 min at 500 *g* to obtain a crude nuclear fraction. The supernatant was centrifuged at 20,000 *g* for 5 min at 4°C. The resulting supernatant was centrifuged at 100,000 *g* for 90 min at 4°C to obtain the cytoplasmic fraction. The pellet of the 100,000 *g* spin was washed three times in homogenization buffer and resuspended in 150 μl of homogenization buffer to obtain the membrane fraction. Protein in each fraction was determined spectrophotometrically (Bradford, 1976). Equal amounts of protein (125 μg) from each fraction were incubated with 4.25 μCi L-[methyl-^{3}H]SAM in a total volume of 50 μl for 1 h at 37°C. The labeling reaction was stopped by adding 12.5 μl of 5× SDS-PAGE sample buffer (0.3 M Tris-HCl, pH 6.8, 45% glycerol, 1.4 M 2-mercaptoethanol, 10% SDS, 0.001% bromophenol blue) and holding the samples in a boiling water bath for 5 min. In vitro–labeled proteins from the cell fractions were then analyzed by 7.5–15% gradient SDS-PAGE as described below.

Gel Electrophoresis

Discontinuous SDS-PAGE (Laemmli, 1970) was performed with 19-cm separating gels composed of polyacrylamide gradients of 6–12, 7.5–15, or 8.5–15%, depending upon the experiment. Gels were fixed, stained with Coomassie blue, and then destained. Gels containing proteins labeled with [^{32}P]orthophosphate were dried and placed in contact with Kodak XAR film to produce an autoradiographic image. Gels containing tritium-labeled proteins were prepared for fluorography by washing the gel for 1 h in three changes of deionized water, followed by immersion in 1 M sodium salicylate for 30 min (Chamberlain, 1979). After drying, the gels were exposed to preflashed Kodak XAR film (Laskey and Mills, 1975) and stored at −70°C (Bonner and Laskey, 1974) with an intensifying screen. Quantitative comparisons of methyl-labeled proteins were obtained by scanning fluorograms into an analysis program (Molecular Analyst; Bio Rad Laboratories, Hercules, CA).

For 2D IEF × SDS-PAGE, PC12 cell proteins were labeled as described above and harvested in a lysis buffer appropriate for IEF (Aletta and Greene, 1987). Equal TCA-precipitable counts per minute of cell lysates were subjected to IEF with pH 5–7, and 3.5–10 ampholines at a ratio of 4:1, respectively. Proteins in IEF gels were further resolved by SDS-PAGE (the second-dimension) using 12-cm separating gels composed of 7.5–15% polyacrylamide gradient. A standard IEF gel containing only lysis solution was used to determine the pH range. Fluorographic images of tritium-labeled proteins were generated as described above.

DNA Methylation

Total methylation of cytosines in DNA was determined by first digesting PC12 cell DNA with MspI, which cleaves methylated or unmethylated CCGG sequences (Bestor et al., 1984). The products are 5′ labeled with [^{32}P]ATP, and digested with nuclease P1. Methyl-cytosines are resolved from unmethylated cytosines by chromatography in isobutyric acid/water/ammonium hydroxide (66:33:1) and detected by autoradiography (Bestor et al., 1984). Liquid scintillation spectrometry was used to quantify the results.

Immunoprecipitation

Tyrosine phosphorylation of Trk in the presence or absence of DHCA was assessed by immunoprecipitation followed by Western blotting. Cells were treated with 100 ng/ml NGF for 5 min, with or without 1 μM DHCA. Cells were then washed three times in ice-cold PBS and harvested in 1 ml of a lysis buffer containing 1% Triton X-100, 150 mM NaCl, 50 mM Tris-HCl, pH 8.0, 25 mM NaF, 5 mM EGTA, 5 mM EDTA, 2 μM PMSF, and 100 U/ml aprotinin. Insoluble material was removed by centrifugation at 4°C for 10 min at 13,000 *g*. Samples were precleared with 6 mg PAS for 2 h followed by centrifugation for 10 min at 13,000 *g*. Anti-phosphotyrosine monoclonal antibody (clone PT-66) was added to equal amounts of lysate protein for 2 h on a rotating platform at 4°C followed by addition of 3 mg of PAS and incubation on the rotating platform at 4°C for 1 h. PAS beads were then recovered by centrifugation at 13,000 *g* for 10 min, and washed three times with 1 ml of 1% Triton X-100 lysis buffer followed by two 1-ml washes in lysis buffer without Triton X-100. The PAS beads were then resuspended in SDS-PAGE sample buffer and held in a boiling water bath for 5 min. The precipitated material was resolved by SDS-PAGE (7.5% acrylamide) followed by transfer to Immobilon P membrane (Millipore Corp., Milford, MA). The blot was probed with 1 μg of rabbit polyclonal Trk antiserum (Santa Cruz Biotechnology, Santa Cruz, CA) for 2 h at room temperature followed by incubation with ^{125}I-donkey anti–rabbit IgG. The dried blot was then exposed to Kodak XAR film at −70°C with an intensifying screen.

SAHH Activity

PC12 cells were plated on 150-mm dishes (2×10^7 cells) in the presence of 100 ng/ml NGF plus 10 nM to 3 μM DHCA, or without added DHCA for 7 d. Cells were then washed three times in ice-cold PBS and harvested by scraping in a potassium phosphate buffer, pH 7.0 (25 mM KH_2PO_4, 25 mM K_2HPO_4, 1 mM dithiothreitol, 0.5 μM PMSF, and 10 μg/ml leupeptin). Cells were homogenized in a Dounce homogenizer, and the cell nuclei and debris were removed by centrifugation at 13,000 *g*. Cytosol was then prepared by centrifugation at 100,000 *g* for 90 min and the protein concentration determined. SAHH activity was determined in the synthesis direction by a TLC method described previously by Hershfield (1979).

Results

NGF-specific Induction of Changes in the Pattern of PC12 Cell Methylated Proteins

Previous studies have implicated regulation of protein methylation in the early events of NGF-mediated signal transduction (Seeley et al., 1984; Kujubu et al., 1993). To examine this possibility in greater detail, protein methylation patterns were assessed after metabolic radiolabeling of cellular proteins with L-[methyl-^{3}H]methionine. Fluorograms of the labeled proteins generated from SDS-PAGE and 2D IEF × SDS-PAGE were used to detect specific changes in protein methylation after NGF treatment. The cellular pool of SAM, the predominant methyl donor in all cells, was radiolabeled using L-[methyl-^{3}H]methionine in control PC12 cells, and PC12 cells treated with either 50 ng/ml NGF, 5 nM EGF or 1 μM insulin. These concentrations and growth factors were chosen based on the biological effects produced by each in PC12 cells. NGF at 50 ng/ml produces the maximum neurite outgrowth response. EGF at 5 nM enhances cell proliferation (Huff et al., 1981) and 1 μM insulin is commonly used to effect cell survival in serum-free media (Rukenstein et al., 1991). We have verified that receptors for all three of these growth factors are present on the PC12 cells used in these studies, by ob-

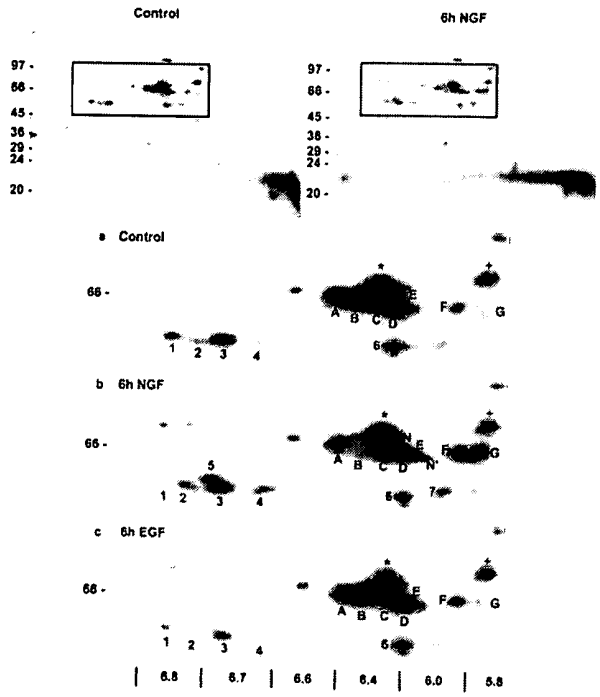


Figure 1. NGF induces specific changes in the pattern of in vivo protein methylation observed by 2D IEF × SDS-PAGE. PC12 cell proteins were metabolically labeled with L-[methyl-^{3}H]methionine (100 μCi/ml) in the presence of anisomycin (10 μM) for 6 h. Whole cell lysates containing equal TCA-precipitable cpm (800,000) were separated by IEF. IEF gels were loaded onto 8.5–15% gradient gels for the separation of methylated proteins by molecular mass. Fluorographic images generated from the control and 6 h NGF (50 ng/ml) conditions are illustrated in the top two panels. Molecular mass standards (in kD) are indicated at the left of each panel. The boxed area in the top panels is enlarged below for control, 6 h NGF and 6 h EGF (5 nM) samples. Fluorograms were generated by exposing x-ray film to the dried gels with an enhancing screen for 10 d at −70°C. The pH gradient of the isoelectric focusing tubes is shown below the enlarged images (*a–c*). Labels (*A*, *, and +) indicate unaffected methylations, while the remaining letters and numbers indicate protein methylations affected by NGF. These results are representative of three independent experiments.

servation of the aforementioned biological effects at the doses indicated (data not shown). Radiolabeled methionine equilibrates with the cellular SAM pool within 20 min (Chelsky et al., 1985). A protein synthesis inhibitor, anisomycin (10 μM), was present during the metabolic radiolabeling to prevent incorporation of L-[methyl-^{3}H]methionine into newly synthesized proteins.

In each of five independent trials, all three trophic factors produced a net increase in methyl group incorporation into PC12 cell protein. After 6 h of treatment, EGF and insulin lead to ~50% greater incorporation (±0.2, SEM) and NGF to a 31% increase (±0.1, SEM). Despite the slightly larger effects of EGF- and insulin-promoted increases in total methyl-^{3}H–labeled protein, NGF treatment consistently yields protein methylation patterns exhibiting more pronounced changes in specific methylated proteins (Fig. 1). Several, indistinctly resolved proteins that migrated between 24 and 20 kD showed increased methyl labeling relative to total protein methylation after 6 h NGF treatment (Fig. 1, *top right*). Among the methylated species detected in control cell lysates, were nine 68–60-kD proteins with isoelectric point (pI) values of 5.8–6.4 (Fig. 1, *A–G*, *, and +). After treatment with 50 ng/ml NGF for 6 h, the proteins labeled *A*, *, and + in Fig. 1 showed little or no reproducible changes in methylation relative to controls. Although there was a net increase in total protein methylation, both increases and decreases in the methylations of specific proteins relative to total protein methylation were detected in response to NGF. Decreased labeling of proteins *B*, *D*, and *E* was detected with NGF (Fig. 1 *b*). Methyl labeling of proteins *C*, *F*, and *G* increased markedly, and methyl group incorporation into two proteins (N and N′) was detected only after 6 h in NGF-treated cells, but not in controls (Fig. 1 *b*).

Significant NGF-induced effects on protein methylation were also detected at pI 6.7–6.8 and 5.8–6.2 in the 50-kD region (Fig. 1, *a* and *b*). Labeling of protein 3 was unaffected by NGF. Decreased methyl labeling of protein 1 was detected after 6 h NGF treatment, while proteins 2, 4, and 7 showed increased methyl labeling, and the [methyl-^{3}H] incorporation into protein 5 was only detected in response to NGF. The effects of NGF on protein methylation are unlikely to be due to incorporation of L-[methyl-^{3}H]methionine into newly synthesized proteins because protein synthesis was inhibited by 97 ± 0.1% (n = 3) under the conditions of these experiments. In addition, no significant increases or decreases in the amounts of the specific proteins analyzed in Fig. 1 were detected by 2D IEF × SDS-PAGE of [^{35}S]methionine-labeled proteins in whole cell lysates, in the absence of protein synthesis inhibition (data not shown).

The NGF-induced changes in methyl labeling of specific proteins were not observed with the two other trophic factors tested. After EGF treatment for 6 h, the methylation pattern of the 68–60-kD proteins was quite different from the NGF-induced changes relative to the control, nonstimulated cells (Fig. 1). The pattern of protein methylations with EGF was more similar to the pattern with control cells, but small increases in the methylation of proteins in this relative molecular weight range were reproducibly detected (B, C, D, and E). NGF, on the other hand, decreased the methylation of proteins B, D, and E. The changes in protein methylations were less marked with EGF than NGF and no methylated protein was detected with EGF that was not detected with controls. Moreover, no significant effects of EGF were detected in the 50-kD set of methylated proteins. Insulin had no significant effects on the methylation of the 68–60- or 50-kD proteins (data not shown).

The effects of NGF on the pattern of protein methylation were also examined by in vitro labeling of proteins in subcellular fractions from NGF-treated cells as an independent approach for detecting the protein methylations that are affected by NGF. In vitro labeling provides a complementary method of analysis that does not require a protein synthesis inhibitor. It also indicates the possible cellular locations of protein methylation. PC12 cells were incubated without or with NGF for 15 min, 6 or 16 h. Nu-

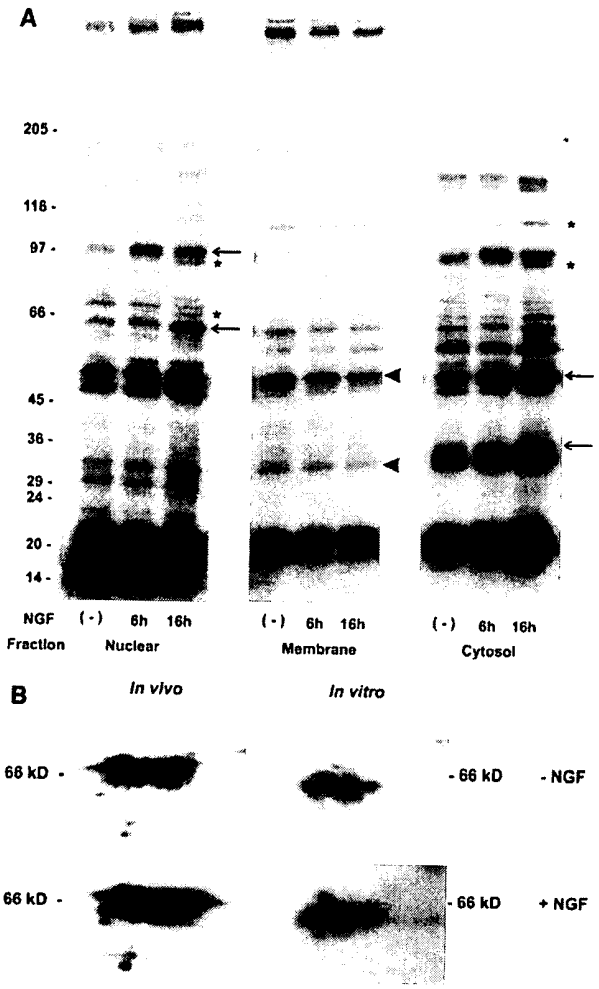


Figure 2. NGF-treatment of intact cells stimulates in vitro protein methylation in cell-free extracts. (*A*) PC12 cells were incubated in the presence or absence of NGF (50 ng/ml) for 6 or 16 h. Proteins in the nuclear (500 *g*, pellet), membrane (100,000 *g*, pellet), or cytosolic (100,000 *g*, supernatant) fractions were radiolabeled with L-[methyl-^{3}H]SAM for 1 h at 37°C. The labeling reaction was quenched by the addition of 5× SDS-PAGE sample buffer followed by holding the reaction tube in a boiling water bath for 5 min. Labeled proteins were separated by 7.5–15% gradient SDS-PAGE. The image shown is a fluorogram of the dried gel. Asterisks indicate methylated proteins not detectable in extracts from non-NGF-treated control cells. Arrows identify increases of 67% or greater. The arrowheads (membrane fraction) point out decreases of >30%. These results were reproduced in an independent experiment. Specific changes at 16 h were as follows: nuclear proteins at 97 and 64 kD increase 2.4-fold; cytosolic proteins at 50 kD and 35 kD increase 67 and 79%, respectively. (*B*) Cell proteins were metabolically radiolabeled followed by preparation of cytosol (*in vivo*) or the cytosol was prepared first, followed by incubation with [methyl-^{3}H]SAM (*in vitro*). A portion of the 2D IEF × SDS-PAGE fluorograms from control (*−NGF*) versus 6 h NGF treatment (*+NGF*) are displayed. There are five individual protein spots migrating at 64 kD, which are most easily discerned in the in vivo +NGF condition. The 64-kD series of spots in each of the other fluorograms (four spots each) are superimposable with those of the in vivo +NGF condition.

clei (500 *g*), membranes (100,000 *g*, pellet), or cytosol (100,000 *g*, supernatant) were isolated and incubated with [methyl-^{3}H]SAM for 1 h at 37°C, in vitro. Methylated proteins were resolved by 7.5–15% gradient gel SDS-PAGE and detected by fluorography. Changes in protein methylation were detected in nuclei, membranes, and cytosol (Fig. 2). Increases in protein methylation were detected after 15 min of NGF treatment (data not shown) and continued to increase from 6 to 16 h (Fig. 2 *A*). Moreover, several of the proteins that showed increases in methylation have similar molecular weights as those detected by 2D IEF × SDS-PAGE (Fig. 1) after metabolic radiolabeling of intact cells (e.g., 68–64 kD, and 50 kD). Thus, changes in the methylation of proteins in response to NGF was confirmed by two independent approaches. In the nuclear fraction, time-dependent increases (more than twofold) in protein methylation in response to NGF treatment were detected in proteins migrating at 97, 94, 67, and 64 kD. In the membrane fraction, the methylation of proteins migrating at 50 and 34 kD decreased after 6 or 16 h of NGF. Several proteins in the cytosolic fraction showed increased methylation. Most prominent among these were 114, 94, 50, and 35 kD. Additional NGF-induced effects were observed by the in vitro method. This is most likely due to the higher specific activity of the [methyl-^{3}H]SAM pool in vitro than in the intact cell experiments and the enrichment of proteins by subcellular fractionation. In addition, the in vitro labeling experiments indicate that both the requisite methyltransferase and protein substrate were present at the time of analysis in each of the subcellular fractions that were isolated from NGF-activated cells. Moreover, irrespective of the specific mechanisms responsible for activation of the methylation of specific proteins by NGF, the state of activation was stable during preparation of the fractions.

To validate this approach further, cytosol was prepared from cells treated with or without NGF after metabolic radiolabeling, and replicate cultures without metabolic radiolabeling were also processed to obtain cytosol for in vitro protein methylation. Comparisons of the 2D IEF × SDS-PAGE fluorograms from each type of preparation indicate that several of the same proteins resolved in the 64 kD range (pI 5.8–6.4) are similarly increased after NGF treatment by either method of analysis (Fig. 2 *B*). The in vitro data obtained after NGF treatment for 6 h confirm that changes in protein methylation triggered by NGF can occur whether or not protein synthesis has been inhibited. Thus, the in vivo and in vitro experiments independently indicate that NGF produces diverse, marked effects on the regulation of the methylation of several specific proteins.

NGF Affects Protein Methylation during Early and Delayed Signaling

The time dependence of NGF-induced changes in protein methylation was examined to determine the onsets and durations of changes in protein methylation. PC12 cells were incubated with NGF for 15 min to 24 h to examine protein methylation during early stages of NGF signaling and for 4 d to examine changes occurring concurrently with the appearance of neurites. Total protein methylation

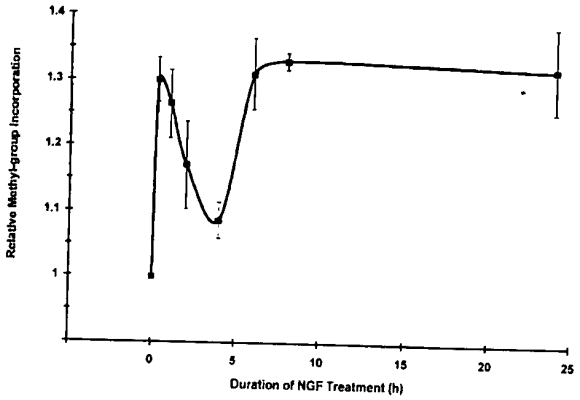


Figure 3. The time dependence of total methyl group incorporation into proteins in response to NGF. PC12 cell proteins were metabolically labeled with L-[methyl-^{3}H]methionine (100 μCi/ml) for 6 h in the presence of anisomycin (10 μM). NGF (50 ng/ml) was added to the cell cultures for the times indicated. Cells were lysed in SDS-PAGE sample buffer. Total methyl labeling of proteins (counts per minute) was measured for each condition after TCA precipitation of an aliquot of the sample. The relative amount of the ^{3}H-labeled methyl groups in lysate proteins at each time point is given with respect to nontreated control cells. The data shown are the means ± SEM for five independent experiments.

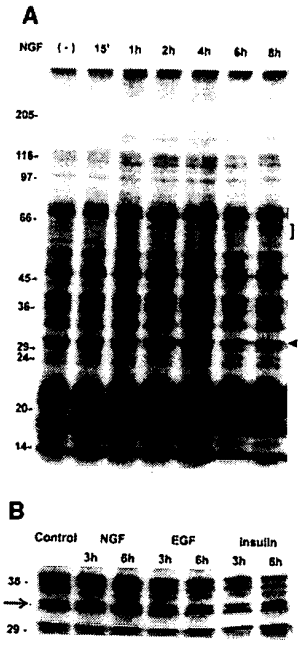


Figure 4. Changes in the pattern of protein methylation are NGF specific and time dependent. (*A*) PC12 cell proteins were metabolically labeled with L-[methyl-^{3}H]methionine (100 μCi/ml) for 6 h in the presence of anisomycin (10 μM) with or without NGF (50 ng/ml) for the times indicated. Whole cell lysates containing equal TCA-precipitable cpm (300,000) were loaded in each lane. Proteins were separated by 7.5–15% gradient SDS-PAGE. The image shown is a fluorogram of the dried gel after exposure to x-ray film for 3 d at −70°C with an intensifying screen. The bracket indicates the migration position of a 64–62-kD complex of proteins that increase in a time-dependent manner. The arrowhead indicates the migration position of an invariant protein band, the density of which does not change by more than a few percent over the time course. By this measure, the 4-h lane is slightly overloaded, and the 6-h lane, underloaded. Nevertheless, the change in the 64–62-kD protein complex is still evident in the latter. The results illustrated here are representative of five independent experiments. (*B*) Metabolically radiolabeled methylated proteins (300,000 cpm) were obtained as described above and loaded on 8.5–15% polyacrylamide gradient gels to more favorably resolve the area between 29 and 36 kD. The image shown is a fluorogram derived from the dried gel after exposure to x-ray film for 2 d at −70°C. The arrow indicates the migration position of a regulated 34-kD protein. Cultures were treated with NGF (50 ng/ml), EGF (5 nM), or insulin (1 μM). The results shown are representative of three independent experiments.

was assessed by the incorporation of [^{3}H]methyl groups into TCA-precipitable protein during NGF treatment for 24 h (Fig. 3). Significant increases in the total amounts of methyl group incorporation were detected within 15 min after the addition of NGF. The response over the early time course is biphasic with an initial rapid burst of methylation that decays between 1 and 4 h of NGF treatment followed by a second, persistent phase of elevated methylation from 6 to 24 h. The pronounced decrease in methyl group incorporation between 1 and 4 h indicates that the effect of NGF on total protein methylation is likely to be complex and dependent upon multiple factors, including methionine uptake, intracellular compartmentalization, and use of the labeled methyl group in other metabolic pathways. Thus, all comparisons of protein methylation patterns examined by gel electrophoresis were standardized relative to total methyl-^{3}H–labeled protein.

Although the resolution of changes in protein methylation in whole cell lysates is reduced on one dimensional SDS-PAGE gels, this method was used to observe the time course of changes in the set of proteins migrating at 64–62 kD. Gradient SDS-PAGE and fluorographic analysis of methyl-^{3}H–labeled proteins in whole cell lysates indicated that there are increases in the methylation of the proteins in this region of the gel, detectable within 1 h after the addition of NGF and persisting for at least 8 h (Fig. 4 *A*). In addition, when using an 8.5–15% acrylamide gradient, a reproducible NGF-specific change in a 34-kD protein is evident at 3 and 6 h of NGF treatment (Fig. 4 *B*), but not at times earlier than 2 h (data not shown).

To determine if any of the changes in methyl group incorporation persist for several days, the labeling studies were repeated in PC12 cells treated with NGF for 4 d. This time point was chosen because ~50% of PC12 cells have neurites ≥20-μm long after 4 d of NGF (Greene et al., 1982). In addition, it was reasoned that some protein methylations may occur during neurite outgrowth that were not detected during early stages of NGF action. Thus, after 4 d of NGF, methylated proteins were analyzed by metabolic radiolabeling of cell proteins followed by 2D IEF × SDS-PAGE (Fig. 5) as described above. The pattern of [^{3}H]methyl incorporation into PC12 cell proteins relative to total methylated protein indicates that many of the changes induced by 6 h of NGF (Fig. 1) persist for several days. Moreover, several of the NGF-induced increases and decreases in protein methylation observed after 6 h of NGF treatment were greater after 4 d of NGF treatment. Methylation of proteins between 68 and 60 kD (C, F, G, N, and N′) and in the 55–50 kD region (2–4, 6, and 7) increased relative to both the control (Fig. 5) and the 6 h NGF result (Fig. 1). Proteins C, F, G, N, and N′ appear to be the same proteins (based on pI and M_r) that showed increases in methylation after 6 h of NGF treatment. Also, after 4 d of NGF treatment, the methylation of proteins B, D, and E, as well as protein 1 in the 50-kD region decreased relative to the control or 6-h NGF result. A

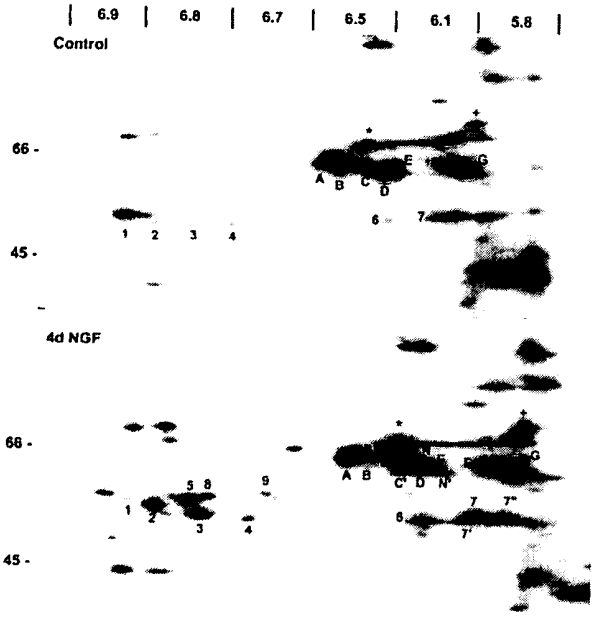


Figure 5. NGF-induced changes in protein methylation persist and progress during prolonged NGF treatment. PC12 cells were incubated with or without NGF (50 ng/ml) for 4 d. Cells were labeled with L-[methyl-^{3}H]methionine (100 μCi/ml) in the presence of anisomycin (10 μM) for 6 h. Equal TCA-precipitable cpm (750,000) of cell lysates were subjected to IEF. IEF gels were loaded on 7.5–15% gradient SDS-PAGE gels to separate methylated proteins by molecular mass. The images shown are fluorograms of the dried gels after exposure of x-ray film for 12 d at −70°C with an intensifying screen. The areas of the gels shown are similar to the boxed areas shown in Fig. 1. The pH gradient of the IEF gels is indicated at the top of the figure. Labels are as in Fig. 1.

number of novel methylated proteins were also observed after 4 d of NGF treatment that were not detected after 6 h. These include proteins B′, C′ in the 60-kD region of the gel, and proteins 7′, 7″, 8, and 9, appearing in the 50-kD region of the gel. The distinctly different time courses for the methylations of specific proteins suggest that NGF may regulate the methylation of proteins involved in multiple signaling pathways. Moreover, most of the NGF-dependent changes in methylated proteins that were detected are not associated with transient phenomena.

Inhibition of Methylation Inhibits Neurite Outgrowth

The marked effects of NGF on the methylation of specific proteins calls attention to the possibility that protein methylation is required for NGF signal transduction and neurite outgrowth. This possibility was examined by observing NGF-mediated neurite outgrowth after inhibiting methylation using two approaches based on the biochemical pathway outlined in Fig. 6. One approach was inhibition of SAHH by a mechanism-based, substrate analogue inhibitor, DHCA (Liu et al., 1992). SAHH catalyzes the reversible hydrolysis of SAHcy to homocysteine and adenosine by an NAD-dependent mechanism (De la Haba and Cantoni, 1959; Liu et al., 1992). The reaction is driven in

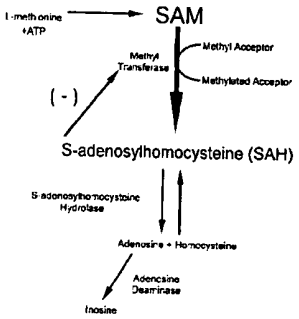


Figure 6. The methylation pathway. SAHH catalyzes the reversible hydrolysis of SAHcy. Inhibition of SAHH by DHCA or homocysteine leads to an accumulation of SAHcy and inhibition of methyltransferases.

the hydrolysis direction in vivo by the adenosine deaminase–catalyzed conversion of adenosine to inosine. DHCA blocks the methylation pathway by inhibiting SAHH. Like the normal substrate, SAHcy, DHCA is oxidized by SAHH coupled to NAD^+ reduction, but the 3-keto DHCA formed cannot be hydrolyzed, thereby trapping SAHH in its inactive, NADH form (Liu et al., 1992). SAHH inhibition results in an accumulation of its substrate, SAHcy, a potent competitive inhibitor of methyltransferases (Hildeshein et al., 1972).

The concentration dependence for inhibition of neurite outgrowth by DHCA (Fig. 7 *A*) was determined by incubating PC12 cells with NGF for 7 d with or without 10 nM to 3 μM DHCA. Neurite outgrowth was scored after 7 d of NGF treatment as described under Materials and Methods. The concentration of DHCA required to inhibit neurite outgrowth by 50% was ~100 nM. A second approach that was used to test for a requirement of protein methylation in neurite outgrowth was inhibition of SAHH by addition of extracellular homocysteine, alone or in combination with DHCA. When intracellular levels of homocysteine rise, the SAHH-catalyzed reaction is driven in the synthesis direction (De la Haba and Cantoni, 1959; Hasobe et al., 1989), i.e., towards the formation of SAHcy (Fig. 6). Excess homocysteine acts cooperatively with inhibitors of SAHH in many other cell types (Chiang et al., 1977; Kredich and Martin, 1977; Backlund et al., 1986; Hasobe et al., 1989), and the combination of the two is expected to elicit more inhibition of methylation than either alone. As shown in Fig. 7 *B*, 100 nM DHCA reduced NGF-induced neurite outgrowth to 59% of control (NGF alone), and 100 μM homocysteine reduced NGF-induced neurite outgrowth to 42%. The combination of 100 nM DHCA plus 100 μM homocysteine inhibited the neurite outgrowth by more than 80%. The findings that homocysteine alone inhibits neurite outgrowth, and that homocysteine enhances the inhibitory effect of a low dose of DHCA on neurite outgrowth indicate that the effect of DHCA on neurite outgrowth is most likely due to inhibition of methylation.

Fig. 8 illustrates the effects of DHCA or DHCA plus homocysteine on cell morphology observed by phase contrast microscopy of NGF-treated PC12 cells. After 7 d of NGF-treatment, PC12 cells flatten, hypertrophy, and elaborate long neurites (Fig. 8 *B*). In the presence of 1 μM DHCA and NGF for 7 d, the cells continue to respond to NGF by flattening, but neurite outgrowth is greatly attenuated (Fig. 8 *C*). Short, spike-like processes are observed.

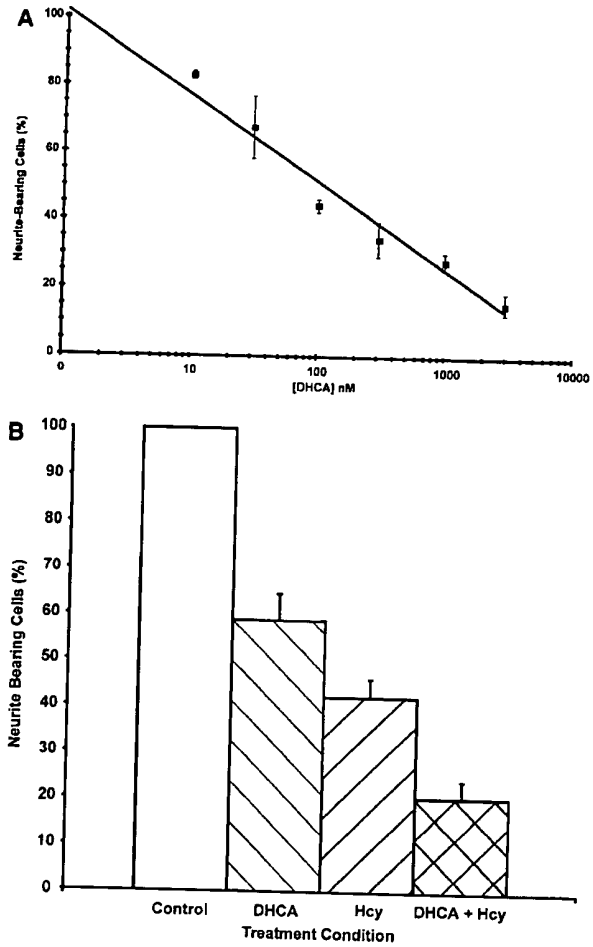


Figure 7. Concentration-dependent inhibition of neurite outgrowth by inhibitors of methylation. (*A*) Effect of DHCA. PC12 cells on 35-mm collagen-coated tissue culture dishes (300,000 cells per dish) were treated with NGF (50 ng/ml), or NGF and DHCA (10 nM to 3 μM) for 7 d. Neurite-bearing cells were scored as described under Materials and Methods. Cultures treated with NGF and no DHCA received the DMSO vehicle only, which had no effect on neurite outgrowth. The line was fit by least squares regression analysis. The data shown are the means ± SD for three independent experiments. (*B*) Effect of homocysteine alone and in combination with DHCA. All cultures were treated with NGF (50 ng/ml) and scored for the presence of neurites after 7 d. The data shown are the means ± SD for three independent experiments. The white bar represents the NGF control condition without DHCA or homocysteine. Results from cultures treated with 100 nM DHCA are represented by the adjacent striped bar (*DHCA*), and those from cultures treated with 100 μM homocysteine in the next striped bar (*Hcy*). The cross-hatched bar indicates results from cultures treated with 100 nM DHCA plus 100 μM homocysteine (*DHCA + Hcy*).

In the presence of 100 nM DHCA plus 100 μM homocysteine, the cells appear more flattened than control cells, but little neurite outgrowth is observed (Fig. 8 *D*). Thus, inhibition of methylation by DHCA, homocysteine, or the cooperative action of the two together suggest that methylation is an important requirement for NGF-induced neurite outgrowth. In addition, the morphological evidence (Fig. 8) indicates that not all NGF actions (e.g., the cell-flattening response) are inhibited by factors that inhibit methylation (see below).

DHCA-induced inhibition of SAHH was verified by direct measurement of SAHH activity. DHCA at 1 nM produced 70% inhibition and >90% inhibition was achieved at 10 nM. The consequence of this inhibition (elevated intracellular level of SAHcy) is expected to inhibit protein methylation (Fig. 6). Thus, the effect of DHCA on protein methylation was directly determined by assessing protein methylation in PC12 cells using metabolic radiolabeling with L-[methyl-^{3}H]methionine in the presence of anisomycin. The dose-dependent inhibition of protein methylation observed is similar to that of the DHCA effect on neurite outgrowth (Fig. 9). DHCA (1 μM) inhibits total methyl group incorporation measured by TCA-precipitation of equal aliquots of whole cell lysates as described under Materials and Methods. The total reduction in radiolabeled, methylated protein was 40% in control, non-NGF-treated cells, and 52% in cells stimulated with NGF for 6 h. Similar differences in the inhibitory effects (±NGF) were observed at all doses of DHCA above 30 nM (data not shown). The inhibitory action of DHCA is thus greater in NGF-treated cells. In conclusion, these experiments indicate that protein methylation is markedly inhibited by DHCA at concentrations that inhibit neurite outgrowth.

In contrast, neither NGF alone nor DHCA with or without NGF have a significant effect on total methylation of cytosines in PC12 cell DNA. The percentage of total cytosines methylated in NGF-treated cells (53.3 ± 1.5) versus that in cells treated with NGF plus DHCA (48.1 ± 0.2) was not statistically significant (n = 3, Student's paired t test; $P > 0.05$, two tails). All measurements were performed after 24-h treatments.

DHCA Treatment Does Not Interfere with Cell Proliferation or Early NGF-mediated Signaling Events

Previous efforts to examine the possible role of methylation in NGF signal transduction (Seely et al., 1984; Haklai et al., 1993; Kujubu et al., 1993) used methylation inhibitors at relatively high concentrations that can produce cytotoxicity (Seeley et al., 1984) and inhibition of receptor tyrosine kinases (Meakin and Shooter, 1991; Maher 1993). The methylation inhibitor used in this study, DHCA, is much less toxic than the related adenosine analogues used previously and is effective at concentrations four orders of magnitude lower than those used previously (Seeley et al., 1984; Meakin and Shooter, 1991; Haklai et al., 1993; Kujubu et al., 1993; Maher, 1993). DHCA does not affect the growth rate of PC12 cells in serum-containing medium (Fig. 10 *A*). In the absence of serum, PC12 cells degenerate via programmed cell death. DHCA has no effect on the rate that cell death is induced (Fig. 10 *A*). In the absence of serum, NGF rescues PC12 cells from programmed cell death (Rukenstein et al., 1991). Also shown in Fig. 10 *A*, DHCA had no effect on the protection afforded by NGF. In addition to confirming the absence of cytotoxicity, these findings suggest that DHCA does not interfere with the ability of NGF to activate its tyrosine kinase receptor (Trk), which prevents programmed cell death

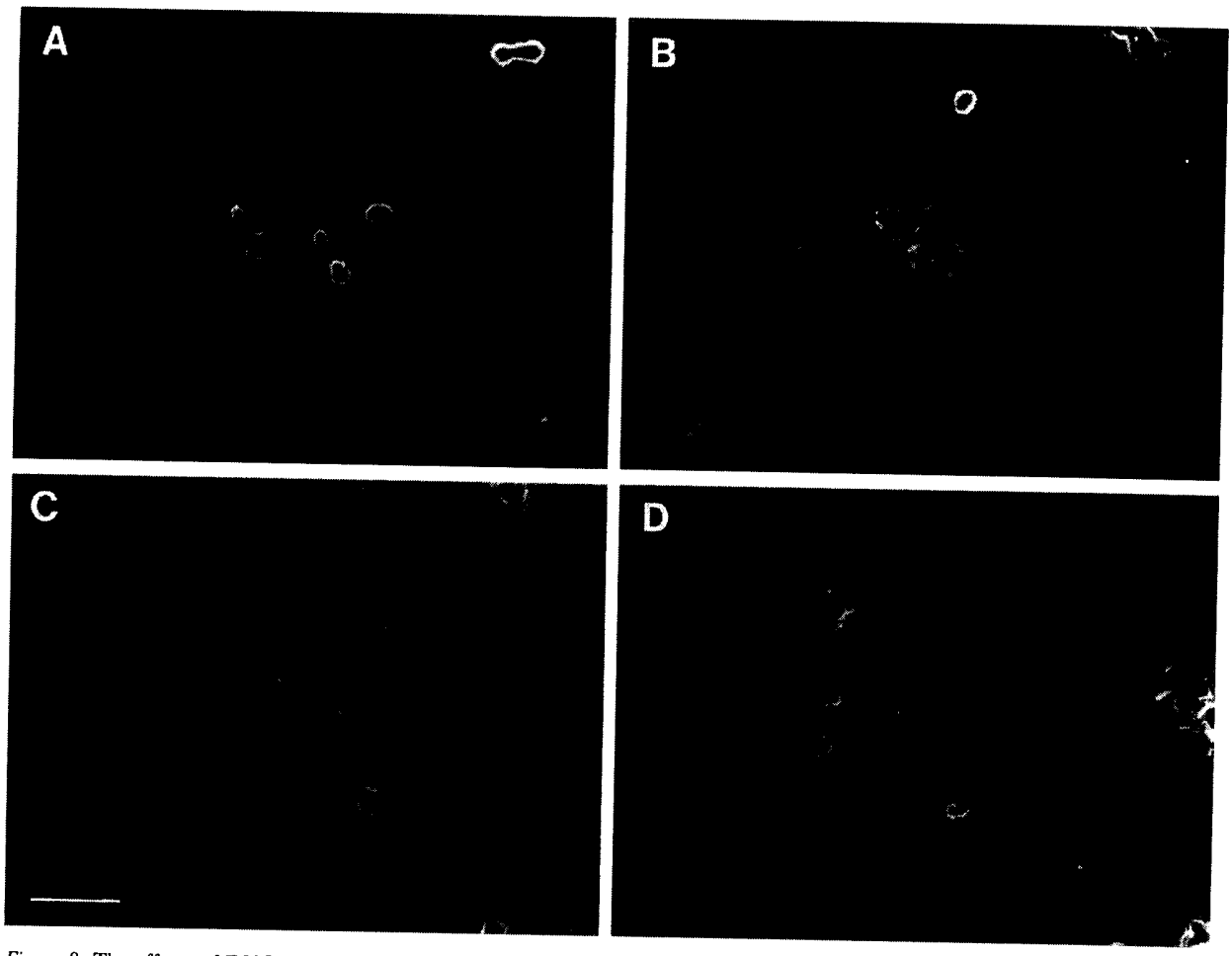


Figure 8. The effects of DHCA and homocysteine on neurite outgrowth and cell morphology. PC12 cells were plated on 35-mm collagen-coated tissue culture dishes at a density of 300,000 cells per dish. Cells were cultured for 7 d under control conditions without NGF (*A*), with 50 ng/ml NGF (*B*), NGF plus 1 μM DHCA (*C*), or NGF with 100 nM DHCA and 100 μM homocysteine (*D*). Bar, 50 μm.

(Kaplan and Stephens, 1994). Thus DHCA has no detectable effects on the growth or survival of PC12 cells under any of the conditions examined.

Since other adenosine-analogues, which are putative methylation inhibitors, are known to inhibit phosphorylation and activation of Trk (Meakin and Shooter, 1991; Maher 1993), the effect of DHCA on the tyrosine phosphorylation of Trk was also determined. PC12 cells were incubated with or without 1 μM DHCA for 1 h, and NGF was then added for 5 min. Immunoprecipitation with a phosphotyrosine antibody was followed by SDS-PAGE and immunoblotting with anti-Trk antibody. DHCA had no detectable effect on the Trk tyrosine phosphorylation mediated by NGF (Fig. 10 *B*), the primary event in NGF signaling.

To determine if DHCA affects other phosphorylation events downstream of Trk activation, total phosphoproteins were analyzed by SDS-PAGE. Several NGF-induced increases in phosphorylation of proteins are resolved by this method (Fig. 10 *C*). The rapid increase in the phosphorylation of tyrosine hydroxylase (Halegoua and Patrick, 1980) that occurs after NGF treatment is not affected by DHCA. The overall phosphorylation pattern was similarly unaffected by DHCA. Two other NGF-induced increases in phosphorylation, which require long-term (7 d) NGF treatment, are temporally associated with neurite outgrowth. Unlike phosphorylation of tyrosine hydroxylase, the increased ^{32}P incorporation into the 64-kD chartin (Aletta and Greene, 1987) and β-tubulin (Aletta, 1996; Black et al., 1986) is diminished by exposure of the cells to DHCA. The inhibitory effect of DHCA on the phosphorylations of these two proteins is the predicted consequence of DHCA inhibition of neurite outgrowth. Thus, DHCA affects delayed phosphorylation events in NGF signaling but does not inhibit phosphorylations associated with the early signaling events studied here.

DHCA Effects on Neurite Outgrowth Are Diminished by prior NGF Treatment and Are Rapidly Reversible

To explore further the functional significance of the time dependence of protein methylation during NGF-mediated neurite outgrowth, DHCA (1 μM) was added to PC12 cells at the same time that NGF was added or 1–6 d after

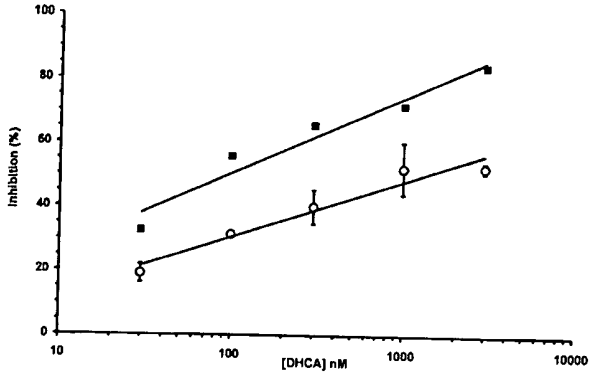
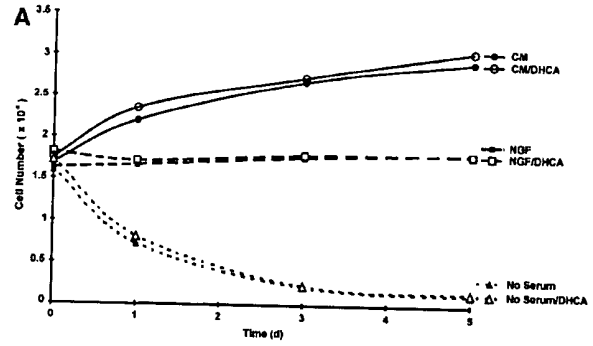


Figure 9. Dose-dependent effect of DHCA on total protein methylation in PC12 cells correlates with the DHCA effect on inhibition of neurite outgrowth. Inhibition of NGF-stimulated neurite outgrowth (■) by DHCA was compared to the inhibition of total protein methylation (○) at the same doses of DHCA. PC12 cells were treated with DHCA (30 nM to 3 μM) and NGF (50 ng/ml) for 6 h in the presence of anisomycin to determine total protein methylation as described under Materials and Methods. Inhibition is plotted relative to control cultures treated with NGF only. The neurite outgrowth data are replotted from Fig. 7 *A*.

NGF addition. Under these conditions, the later DHCA was added after NGF, the less its inhibitory effect (Fig. 11 *A*). The decrease in inhibition showed a linear dependence on the time at which DHCA was added to the cells. Once the NGF-signaling cascade begins, neurite outgrowth is progressively less sensitive to inhibition of methylation. This would occur if the rates of demethylation were low for those methylated proteins that are required for neurite outgrowth. Alternatively, neurite outgrowth may be insensitive to inhibiting methylation once neurite outgrowth proceeds past a critical, committed step. While initiation of neurite outgrowth from PC12 cells after exposure to NGF is slow (requiring 7 d to achieve ~100% neurite bearing cells), PC12 cells can rapidly regenerate after preformed neurites are disrupted as described under Materials and Methods. Fig. 11 *B* shows the concentration dependence of inhibition of neurite regeneration by DHCA. Neurite regeneration was inhibited 40% by 3 μM DHCA. In contrast, if 3 μM DHCA is used when NGF is first added to the cells, it inhibits neurite outgrowth by 84% (Fig. 7 *A*). As described above, the difference in the sensitivities of regeneration and de novo neurite outgrowth to inhibition by DHCA is consistent with slow rates of demethylation or decreased sensitivity to DHCA once differentiation has reached a critical stage.

The onset of neurite outgrowth after DHCA is removed was examined to compare the rate of reversing the effect of DHCA on neurite outgrowth to the rate of NGF-induced neurite outgrowth when DHCA had never been present. In addition, removal of DHCA was used to detect protein methylations that are associated with neurite outgrowth during the reversal period. Reversibility of the effects of DHCA on neurite outgrowth and protein methylation was assessed by incubating PC12 cells with NGF and DHCA for 10 d, followed by removal of the DHCA. Interestingly, the inhibitory effect of DHCA on neurite out-



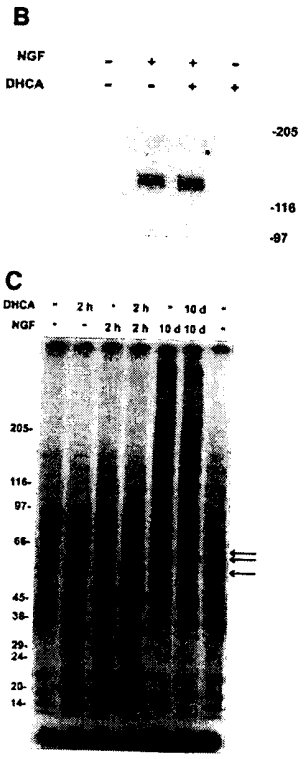


Figure 10. DHCA treatment does not interfere with cell growth, NGF-mediated survival, or rapid onset NGF-stimulated phosphorylations. (*A*) PC12 cells were plated on 24-well collagen-coated dishes at a density of 100,000 cells per well. Cells were allowed to attach to the dishes overnight, and then were washed with RPMI-1640 medium to remove serum. Cells were then cultured in complete (serum-containing) medium (*CM*, ○ and ●), plus or minus 1 μM DHCA; RPMI-1640 without serum plus 50 ng/ml NGF, plus or minus 1 μM DHCA (*NGF*, □ and ■); RPMI-1640 medium without serum or NGF, plus or minus 1 μM DHCA (*No serum*, △ and ▲). Similar results were observed in two other independent experiments. (*B*) Effect on Trk phosphorylation. PC12 cells were treated with or without NGF (50 ng/ml) for 5 min. Where indicated, 1 μM DHCA was added to cultures for 1 h before treatment with NGF. Cell lysates were equalized for total protein, immunoprecipitated with anti-phosphotyrosine antibody (as described under Materials and Methods), resolved by 7.5% SDS-PAGE and transferred to Immobilon P membranes. The blot was probed with anti-Trk antibody, followed by donkey anti-rabbit ^{125}I-IgG. The image shown is an autoradiogram of the Western blot after a 5-d exposure at −80°C. This result was verified in a second independent experiment. (*C*) PC12 cells were treated with or without NGF (50 ng/ml) for 2 h. DHCA (1 μM) was added 1 h before the addition of NGF where indicated. Cells were radiolabeled with [^{32}P]orthophosphate for 2 h as described under Materials and Methods. For long term NGF-induced phosphorylations, PC12 cells were grown in the presence of NGF (50 ng/ml) plus or minus 1 μM DHCA for 10 d. Cells were then radiolabeled with [^{32}P]orthophosphate for 2 h. For both short and long term NGF treatment experiments, cell lysates were equalized for total TCA-precipitable ^{32}P-labeled protein and separated on 6–12% gradient SDS-PAGE. The image shown is an autoradiogram of the dried gel. Exposure time was 16 h. The arrows at the right indicate, in decreasing M_r, the positions of the previously identified phosphoproteins, 64-kD chartin, 60-kD tyrosine hydroxylase, and 55-kD β-tubulin.

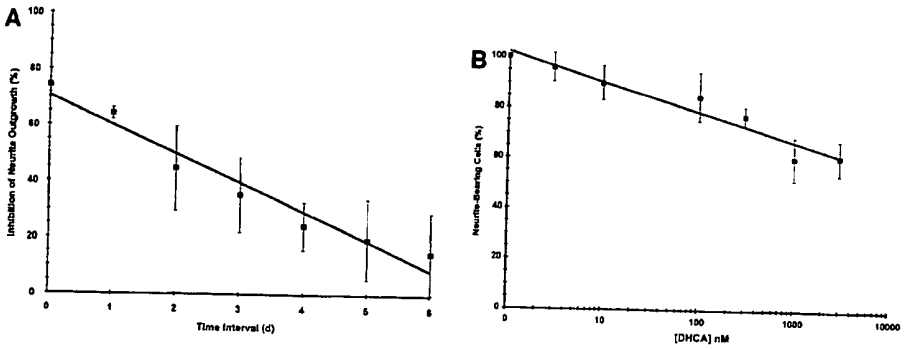


Figure 11. The effect of prior NGF treatment on the neurite inhibiting action of DHCA. (*A*) The effect of varying the time interval between NGF treatment and addition of DHCA. PC12 cells were treated with NGF (50 ng/ml) for 7 d. DHCA (1 μM) was added to the cells either at the time NGF was initially added to the cell cultures (time zero), or after a time interval of 1, 2, 3, 4, 5, or 6 d. Cells were scored for the presence of neurites after 7 d of NGF. The data shown are the means ± SD for three independent experiments. (*B*) The effect of inhibition of methylation on neurite regeneration. PC12 cells were grown in NGF (50 ng/ml) for 7–14 d. NGF was removed from the cultures by an extensive washing procedure described under Materials and Methods. Cells were removed from the culture dishes by trituration and replated without or with DHCA (3 nM to 3 μM). NGF was added to all cultures 1 h after plating. Neurite regeneration was scored 16–24 h later. The line was fit by least squares regression analysis. The data shown are the means ± SD for three independent experiments.

growth was rapidly reversed (Fig. 12 *A*). No lag was observed before an increase in neurite outgrowth was detected. The $t_{1/2}$ for reversal was 12 h (Fig. 12 *A*). The rate of neurite outgrowth when DHCA is removed is, therefore, about eightfold faster than the normal $t_{1/2}$ for neurite outgrowth when NGF is added to naive cells. Thus, the NGF-dependent signaling events that prime the cells (Burstein and Greene, 1978; Greene et al., 1982) for neurite outgrowth were not inhibited greatly by the methylation inhibitor, DHCA.

Upon removal of DHCA, the rapid onset of neurite outgrowth was paralleled by rapid reversal of its inhibitory effects on the methylation of several proteins. Proteins that increase by 25% or more in methylation upon removal of DHCA are indicated by the arrows, brackets, and asterisks shown in Fig. 12 *B*. The molecular weights of proteins that showed increased methylation migrate at 150, 120, 94, 70/68, 50, 48/47, 40, and 35 kD. The time dependencies for methylation were different for different proteins. These results, together with the time course (Figs. 4 and 5), inhibitor studies (Figs. 7–10), and specificity of the NGF-induced effects on protein methylation (Fig. 1) indicate that regulation of protein methylation is required for neurite outgrowth.

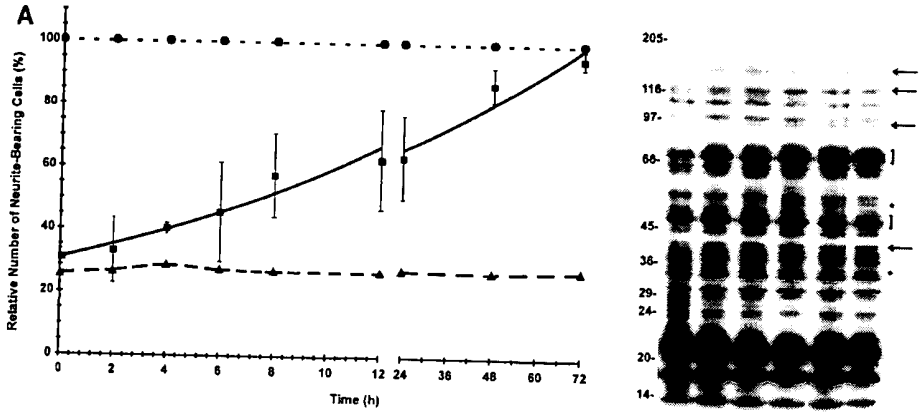


Figure 12. Reversal of DHCA-induced inhibition of methyltransferases results in rapid neurite outgrowth and a time-dependent increase in protein methylation. (*A*) PC12 cells were plated on 35-mm collagen-coated tissue culture dishes at a density of 300,000 cells per dish in NGF (50 ng/ml) plus or minus 1 μM DHCA for 7 d. The medium was then changed repeatedly to remove DHCA, and neurite outgrowth was scored and plotted over a 3-d period. Controls were maintained in DHCA/NGF (▲) and NGF alone (●). The experimental condition was treatment with DHCA and NGF for 7 d, followed by NGF alone for 3 d (■). The data are normalized to 100% relative to the number of neurite bearing cells present in cultures treated with NGF alone. The data shown are the means ± SD for three independent experiments. The line shown for the reversal data is the best fit to a single exponential function with a $t_{1/2}$ of 12 h. After DHCA removal, the protein methylation pattern was also assessed by SDS-PAGE. (*B*) Cells were harvested after 10 d of DHCA/NGF treatment or after 7 d DHCA/NGF treatment followed by progressively longer times (*-h*) in the absence of DHCA. Cell lysates were equalized for total TCA-precipitable ^3H counts per minute and separated on 7.5–15% gradient SDS-PAGE. The figure shown is a fluorogram of the dried gel. Exposure time was 5 d. The arrows and brackets indicate the positions of proteins that increased in methylation by 25% or more relative to the 10 d DHCA lane. Asterisks mark the position of protein bands not detected in the 10-d DHCA lane. The changes noted during the reversal period were replicated in four independent experiments.

Discussion

The Subset of Methylated Proteins that Are NGF-regulated

This work extends previous studies (Seeley et al., 1984; Kujubu et al., 1993) of the role of methylation in growth factor signaling mechanisms and further demonstrates the feasibility of exploring this role by metabolic radiolabeling of specific proteins in intact cells and by in vitro labeling of proteins in cell fractions. The results indicate that protein methylation is involved in the NGF-induced signal transduction that leads to neurite outgrowth. Methylation of numerous proteins was detected by metabolic radiolabeling indicating that protein methylation is an ongoing, active process. NGF specifically affects the methylation of a subset of these proteins. Further evidence for a significant role for protein methylation in neurite outgrowth is the marked concurrent inhibitory effects of DHCA on protein methylation and neurite outgrowth. Moreover, removal of DHCA rapidly leads to concomitant increases in protein methylation and neurite outgrowth. The identities and functions of the NGF-regulated proteins remain to be determined. It is likely that some of the proteins identified by this approach represent previously unidentified players in NGF signal transduction. This is significant because specific protein modifications that are temporally coincident with neuronal differentiation may be useful in the elucidation of the mechanism of action of NGF (Chao, 1992; Kaplan and Stephens, 1994). Thus, signal transduction pathways may be regulated in part by regulating the methylation of specific proteins in much the same manner as protein phosphorylation regulates signaling pathways.

Potential Actions of Protein Methylation During NGF Signal Transduction

The data presented here focus on NGF actions on protein methylations rather than methylation of phospholipids, RNA or DNA. Methylation of genes serves as a general signal for repression of transcription (Razin and Cedar, 1991). Hypomethylation of the genome is associated with cellular differentiation in vitro (Bestor et al., 1984). There were no significant effects of NGF and/or DHCA on total DNA cytosine methylation in PC12 cells. Thus, the potent inhibitory action of DHCA on neurite outgrowth is unlikely to be due to an effect on DNA methylation. Previous work also indicates that NGF does not detectably affect phospholipid methylation (Ferrari and Greene, 1985). These authors further demonstrated that complete inhibition of phospholipid methylation is achieved at concentrations of deaza-adenosine and homocysteine thiolactone that do not block neurite outgrowth.

There are several potential mechanisms through which NGF-induced changes in protein methylation may affect biological responses. Gene regulation may be affected, particularly if the modified proteins are involved in gene expression. For example, many heterogeneous nuclear RNPs that associate with pre-mRNAs are methylated on arginine residues (Liu and Dreyfuss, 1995). Modulation of RNA splicing has been proposed as one potential target for protein methylation in signal transduction (Lin et al., 1996). Consistent with this view, a protein arginine methyltransferase activity appears to be constitutively associated with the type I interferon receptor (Abramovich et al., 1997). A role for protein methylation in cell signaling is further substantiated by the finding that methylation of the γ subunit of a heterotrimeric G-protein enhances activation of PI-specific phospholipase C and PI3-kinase in vitro (Parish et al., 1995). If this biochemical effect occurs in intact cells, protein methylation could play an important general role in the amplification of intracellular signals. Protein methylation may also be involved in altering the subcellular localization of specific proteins, as recently suggested for the cell cycle–dependent methylation of protein phosphatase 2A (Turowski et al., 1995).

Neurite outgrowth requires a fundamental reorganization of the cytoskeleton including formation of intermediate filaments and parallel bundles of microtubules oriented longitudinally in neurites, as well as specialized actin microfilament arrays within motile growth cones. Inhibition of neurite outgrowth by DHCA may be related to inhibition of protein methylation, which affects the cytoskeleton. For example, Ras and Rho proteins, modified by isoprenylation and methylation, are involved in redistribution of actin microfilaments during changes in cell shape (Vojtek and Cooper, 1995). Recently, Luo et al. (1996) presented evidence that constitutively active Rac1 in transgenic mice interferes with the normal formation of Purkinje cell axons and dendrites. Since only the methylation step of small G-protein processing is reversible (Rando, 1996), methylation is more likely than other processing steps to be subject to physiological regulation. NGF-induced increased methyl labeling of G-proteins in PC12 cells (Haklai et al., 1993; Kujubu et al., 1993) and the specific proteins observed in this study indicate that NGF may serve as one mediator of regulatory control in responsive neurons. Furthermore, Klein and colleagues have raised the possibility that protein methylation plays an important role in neurulation (Coelho and Klein, 1990; Moephuli et al., 1997) and neurite outgrowth in rat embryo cultures (Coelho and Klein, 1990). The present work lends additional support to this possibility, especially with regard to neurite outgrowth.

DHCA Is a Potent and Nontoxic Inhibitor of Protein Methylation

DHCA is a promising tool for investigating the possible roles of methylation in signal transduction mechanisms. This inhibitor was developed by Liu and colleagues to obtain a more specific, less toxic inhibitor of SAHH than previously used inhibitors (Liu et al., 1992). The 50% effective dosage (ED_{50}) for the effect of DHCA on neurite outgrowth for PC12 cells ($\sim$100 nM) is four orders of magnitude lower than previously used inhibitors (Seeley et al., 1984). Inhibition of neurite outgrowth by $>\sim$85% is difficult to achieve, even at the highest concentration of the inhibitor. Pretreatment of cells with DHCA for several weeks before NGF stimulation, however, does produce nearly complete inhibition even at 7 d of NGF (Cimato, T.R., M.J. Ettinger, and J.M. Aletta, unpublished observations). These observations provisionally indicate that basal, ongoing protein methylation may serve an auxiliary role in neurite outgrowth in addition to the NGF effects

on protein methylation. As expected, DHCA potently inhibits SAHH activity (Hasobe et al., 1989) and consequently leads to decreased total protein methylation. The dose-dependent inhibition of protein methylation observed is quite similar to that for inhibition of neurite outgrowth (Fig. 9). The cooperative effects of homocysteine and DHCA on neurite outgrowth further substantiate a specific effect on methylation. The finding that total protein methylation was inhibited more when NGF was present than when absent, suggests that under the influence of NGF, many of the potential methyltransferases involved in the NGF response may be more sensitive to inhibition due to increased levels of SAHcy. This may occur because, as a result of increased total protein methylation (Fig. 3), intracellular levels of SAHcy are expected to be elevated even before DHCA treatment. NGF has no measureable effect on SAHH activity (Cimato, T.R., M.J. Ettinger, and J.M. Aletta, unpublished results).

DHCA is unique among methylation inhibitors because of its lack of effect on cell flattening, phosphorylation of tyrosine hydroxylase, and tyrosine phosphorylation of Trk. Less specific methylation inhibitors, used to study the role of methylation in PC12 cells previously (Seeley et al., 1986; Kujubu et al., 1993; Maher, 1993; Meakin and Shooter, 1995), blocked each of these actions. The most remarkable property of DHCA in comparison to previously used inhibitors is that DHCA inhibits neurite outgrowth without inhibiting the NGF priming that prepares PC12 cells for neurite outgrowth (Greene et al., 1982). This is best illustrated by the rapid reversal of DHCA effects on neurite outgrowth, which is in marked contrast to that observed with other methylation inhibitors. Reversal of inhibition by 5′-deoxy-S-methyl adenosine requires the same 7 d as required by control PC12 cells when first exposed to NGF (Seeley et al., 1986). Thus, DHCA is clearly more specific than previously used methylation inhibitors. DHCA specifically affects neurite outgrowth without affecting many other signaling effects of NGF, including those associated with priming.

Protein Methylation in Early and Delayed NGF Signaling

Growth factors in general, promote biological responses that are either extremely rapid (within minutes) or require many hours or days to develop fully. The diversity of signaling networks (Pawson, 1985; Hunter, 1995) and the propensity of specific pathways to transmit and/or sustain a given signal (Marshall, 1995) may help to explain the different temporal characteristics of responses. The results of this study and other recent work (Metz et al., 1993; Philips et al., 1993, 1995) indicate that posttranslational modification of proteins by methylation offers another potential mechanism for regulating both early and delayed growth factor signaling.

The importance of early and delayed responses to NGF in PC12 cells, dorsal root ganglion, and sympathetic neurons is well documented (Greene, 1984). One of the most obvious examples of a delayed NGF response is neurite outgrowth. Little neurite outgrowth is evident in PC12 cells treated with NGF for 12–24 h. 50% of the cells extend neurites within 4 d, and nearly all have neurites by 7 d. NGF-mediated neurite outgrowth requires transcription (Burstein and Greene, 1978; Greene et al., 1982). The long latency is apparently due to the progressive accumulation of specific proteins required for neurite assembly during the priming period. This accumulation may depend on protein synthesis (Burstein and Greene, 1978) and/or progressive posttranslational modifications of specific proteins (e.g., as shown here for the methylation of specific proteins). Since regeneration of neurites from PC12 cells previously treated with NGF occurs much more rapidly (within 24 h) than initiation (7 d); the longer time required for neurite outgrowth in the latter is thought to be due to the "priming" events.

When interpreted in light of the priming model of NGF-induced neurite outgrowth, the experiments that used DHCA provide important insights regarding the temporal nature of protein methylation involved in this NGF response. Inhibition of protein methylation by DHCA is most effective in reducing neurite formation when the DHCA is present concomitantly with NGF treatment. There is decreased inhibition of neurite outgrowth as the time interval between NGF treatment and the addition of DHCA increases (Fig. 11 *A*). DHCA is also less effective at inhibiting neurite regeneration (Fig. 11 *B*). These results imply progressive accumulation of the methylated protein product(s) necessary for neurite outgrowth, as was detected at 4 d of NGF treatment (Fig. 5). Slow turnover of methylated proteins is consistent with the reduced effect of DHCA when added after NGF signaling has been initiated.

The time dependencies of specific protein methylations, the effects of DHCA on protein methylation, and the lack of effects of DHCA on early protein phosphorylations and priming are consistent with little or no interdependence between protein methylation signaling events and early, priming events. Moreover, the data presented on the phosphorylation of Trk and other proteins that lie downstream of Trk activation (Fig. 10, *B* and *C*) further substantiate the view that some cellular signaling events are more susceptible to interference from inhibition of protein methylation than others. Autophosphorylation of Trk and the rapid increase in the phosphorylation of tyrosine hydroxylase (Halegoua and Patrick, 1980) are not affected by DHCA. Two other NGF-induced increases in phosphorylation, which require long term (≥7 d) NGF treatment for maximal phosphorylation, are temporally associated with neurite outgrowth. In accord with DHCA-mediated inhibition of neurite outgrowth, but unlike the rapid, early phosphorylation events, the increased ^{32}P incorporation into the 64-kD chartin (Aletta and Greene, 1987) and β-tubulin (Black et al., 1986; Aletta, 1996) is diminished by exposure of the cells to DHCA.

It can thus be concluded that DHCA does not disrupt all early NGF signaling events and priming actions, nor does it abrogate many of the biological effects of NGF signaling (e.g., Fig. 10 *A*, survival; Fig. 9, cell flattening). The effects of DHCA on neurite outgrowth and inhibition of the phosphorylation of neurite-associated proteins (β-tubulin and chartin) are consistent with a selective inhibitory action on an NGF pathway responsible for neurite outgrowth. Based on all of the above results, the protein methylation required is likely to occur downstream from,

or parallel to, rapid NGF actions and priming events. This interpretation does not exclude the possibility that protein methylation of rapid onset also plays an important role in neurite outgrowth. Once priming is achieved, NGF stimulation of some of the rapid protein methylations observed in this study may assume added significance, particularly if they occur at the growth cone.

We are grateful to R. Borchardt for synthesizing the DHCA that was used for these studies.

This work was supported in part by USDA grant No. 9500773, and by the Multidisciplinary Research Pilot Program from the University at Buffalo, State University of New York.

Received for publication 20 March 1997 and in revised form 17 June 1997.

References

Abramovich, C., B. Yakobson, J. Chebath, and M. Revel. 1997. A protein-arginine methyltransferase binds to the intracytoplasmic domain of the IFNAR1 chain in the type I interferon receptor. *EMBO (Eur. Mol. Biol. Organ.) J.* 16:260–266.

Aletta, J.M. 1996. Phosphorylation of type III β-tubulin in PC12 cell neurites during NGF-induced process outgrowth. *J. Neurobiol.* 31:461–475.

Aletta, J.M., and L.A. Greene. 1987. Sequential phosphorylation of chartin microtubule-associated proteins is regulated by the presence of microtubules. *J. Cell Biol.* 105:277–290.

Backlund, P.S., Jr., D. Carotti, and G.L. Cantoni. 1986. Effects of the S-adenosylhomocysteine hydrolase inhibitors 3-deazaadenosine and 3-deazaaristeromycin on RNA methylation and synthesis. *Eur. J. Biochem.* 160:245–251.

Bestor, T.H., S.B. Hellewell, and V.M. Ingram 1984. Differentiation of two mouse cell lines is associated with hypomethylation of their genomes. *Mol. Cell. Biol.* 4:1800–1806.

Black, M.M., J.M. Aletta, and L.A. Greene. 1986. Regulation of microtubule composition and stability during nerve growth factor-promoted neurite outgrowth. *J. Cell Biol.* 103:545–557.

Bonner, W.M., and R.A. Laskey. 1974. A film detection method for tritium-labeled proteins and nucleic acids in polyacrylamide gels. *Eur. J. Biochem.* 46:83–88.

Bradford, M.M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72:248–254.

Burstein, D.E., and L.A. Greene. 1978. Evidence for RNA synthesis-dependent and independent pathways in stimulation of neurite outgrowth by nerve growth factor. *Proc. Natl. Acad. Sci. USA.* 75:6059–6063.

Chamberlain, J.P. 1979. Fluorographic detection of radioactivity in polyacrylamide gels with the water-soluble fluor, sodium salicylate. *Anal. Biochem.* 98:132–135.

Chao, M.V. 1992. Growth factor signaling: where is the specificity? *Cell.* 68: 995–997.

Chelsky, D., B. Ruskin, and D.E. Koshland. 1985. Methyl-esterified proteins in a mammalian cell line. *Biochemistry.* 24:6651–6658.

Chiang, P.K., H.H. Richards, and G.L. Cantoni. 1977. S-adenosyl-L-homocysteine hydrolase: analogues of S-adenosyl-L-homocysteine as potential inhibitors. *Mol. Pharmacol.* 13:939–947.

Coelho, C.N.D., and N.W. Klein. 1990. Methionine and neural tube closure in cultured rat embryos: morphological and biochemical analyses. *Teratology.* 42:437–451.

De la Haba, G., and G.L. Cantoni. 1959. The enzymatic synthesis of S-adenosyl-L-homocysteine from adenosine and homocysteine. *J. Biol. Chem.* 234: 603–608.

Favre, B., S. Zolnierowicz, P. Turowski, and B.A. Hemmings. 1994. The catalytic subunit of protein phosphatase 2A is carboxyl-methylated in vivo. *J. Biol. Chem.* 269:16311–16317.

Ferrari, G., and L.A. Greene. 1985. Does phospholipid methylation play a role in the primary mechanism of action of nerve growth factor? *J. Neurochem.* 45:853–859.

Greene, L.A. 1984. The importance of both early and delayed responses in the biological actions of nerve growth factor. *Trends Neurosci.*13:91–94.

Greene, L.A., and A.S. Tischler. 1976. Establishment of a noradrenergic clonal line of rat adrenal pheochromocytoma cells which respond to nerve growth factor. *Proc. Natl. Acad. Sci. USA.* 73:2424–2428.

Greene, L.A, D.E. Burstein, and M.M. Black. 1982. The role of transcription-dependent priming in nerve growth factor promoted neurite outgrowth. *Dev. Biol.* 91:305–316.

Greene, L.A., J.M. Aletta, A. Rukenstein, and S.H. Green. 1987. PC12 pheochromocytoma cells: culture, nerve growth factor treatment, and experimental exploitation. *Methods Enzymol.* 147:207–216.

Haklai, R., S. Lerner, and Y. Kloog. 1993. Nerve growth factor induces a succession of increases in isoprenylated methylated small GTP-binding proteins of PC12-pheochromocytoma cells. *Neuropeptides.* 24:11–25.

Halegoua, S., and J. Patrick. 1980. Nerve growth factor mediates phosphorylation of specific proteins. *Cell.* 22:571–581.

Hasobe, M., J.G. McKee, H. Ishii, M. Cools, R.T. Borchardt, and E. DeClercq. 1989. Elucidation of the mechanism by which homocysteine potentiates the anti-vaccinia virus effects of the S-adenosylhomocysteine hydrolase inhibitor 9-(trans-2′, trans-3′-dihydroxycyclopent-4′-enyl) derivatives of adenine and 3-deazaadenine on the metabolism of S-adenosylhomocysteine in mouse L929 cells. *Mol. Pharmacol.* 36:490–496.

Hershfield, M.S. 1979. Apparent suicide inactivation of human lymphoblast S-adenosylhomocysteine hydrolase by 2′-deoxyadenosine and adenosine arabinoside. *J. Biol. Chem.* 254:22–25.

Hildeshein, J., R. Hildeshein, and E. Lederer. 1972. New syntheses of S-adenosyl homocysteine analogues, potential methyltransferase inhibitors. *Biochimie (Paris).* 54:989–995.

Hrycyna, C.A., and S. Clarke. 1993. Modification of eukaryotic signaling proteins by COOH-terminal methylation reactions. *Pharmacol. Ther.* 59:281–300.

Huff, K., D. End, and G. Guroff. 1981. Nerve growth factor-induced alteration in the response of PC12 pheochromocytoma cells to epidermal growth factor. *J. Cell Biol.* 88:189–198.

Hunter, T. 1995. Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling. *Cell.* 80:225–236.

Kaplan, D.R., and R.M. Stephens. 1994. Neurotrophin signal transduction by the Trk receptor. *J. Neurobiol.* 25:1404–1417.

Kredich, N.M., and D.W. Martin, Jr. 1977. Role of S-adenosylhomocysteine in adenosine mediated toxicity in cultured mouse T lymphoma cells. *Cell.* 12: 931–938.

Kujubu, D.A., J.B. Stimmel, R.E. Law, H.R. Herschman, and S. Clarke. 1993. Early responses of PC12 cell to NGF and EGF: effect of K252a and 5′-methylthioadenosine on gene expression and membrane protein methylation. *J. Neurosci. Res.* 36:58–65.

Laemmli, U.K. 1970. Cleavage of structural proteins during assembly of the head of bacteriophage T4. *Nature (Lond.).* 222:680–685.

Laskey, R.A., and A.D. Mills. 1975. Quantitative film detection of ^{3}H and ^{14}C in polyacrylamide gels by fluorography. *Eur. J. Biochem.* 56:335–341.

Lee, J., and J. Stock. 1993. Protein phosphatase 2A catalytic subunit is methyl-esterified at its carboxyl terminus by a novel methyltransferase. *J. Biol. Chem.* 268:19192–19195.

Lee, J., Y. Chen, T. Tolstykh, and J. Stock. 1996. A specific protein carboxyl methylesterase that demethylates phosphoprotein phosphatase 2A in bovine brain. *Proc. Natl. Acad. Sci. USA.* 93:6043–6047.

Lin, W.-J., J.D. Gary, M.C. Yang, S. Clarke, and H.R. Herschman. 1996. The mammalian immediate-early gene TIS21 protein and the leukemia-associated BTG1 protein interact with a protein-arginine *N*-methyltransferase. *J. Biol. Chem.* 271:15034–15044.

Liu, Q., and G. Dreyfuss. 1995. In vivo and in vitro arginine methylation of RNA-binding proteins. *Mol. Cell. Biol.* 15:2800–2808.

Liu, S., M.S. Wolfe, and R.T. Borchardt. 1992. Rational approaches to the design of antiviral agents based on S-adenosyl-L-homocysteine hydrolase as a molecular target. *Antivir. Res.* 19:247–265.

Luo, L., T.K. Hensch, L. Ackerman, S. Barbel, L.Y. Jan, and Y.N. Jan. 1996. Differential effects of the Rac GTPase on Purkinje cell axons and dendritic trunks and spines. *Nature (Lond.).* 379:837–840.

Maher, P.A. 1993. Inhibition of the tyrosine kinase activity of fibroblast growth factor receptor by the methyltransferase inhibitor 5′-methylthioadenosine. *J. Biol. Chem.* 268:4244–4249.

Marshall, C.J. 1995. Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation. *Cell.* 80: 179–185.

Meakin, S.O., and E.M. Shooter. 1991. Tyrosine kinase activity coupled to the high-affinity nerve growth factor-receptor complex. *Proc. Natl. Acad. Sci. USA.* 88:5862–5866.

Metz, S.A., M.E. Rubaglia, J.B. Stock, and A. Kowluru. 1993. Modulation of insulin secretion from normal rat islets by inhibitors of the post-translational modifications of GTP binding proteins. *Biochem. J.* 295:31–40.

Miller, M.W., D.M.J. Duhl, B.M. Winkes, F. Arredondo-Vega, P.J. Saxon, G.L. Wolff, C.J. Epstein, M.S. Hershfield, and G.S. Barsh. 1994. The mouse *lethal nonagouti* (a^{x}) mutation deletes the *S-adenosylhomocysteine hydrolase* (*Ahcy*) gene. *EMBO (Eur. Mol. Biol. Organ.) J.* 13:1806–1816.

Mobley, W.C., A. Schenker, and E.M. Shooter. 1976. Characterization and isolation of proteolytically modified nerve growth factor. *Biochemistry.* 15: 5543–5552.

Moephuli, S.R., N.W. Klein, M.T. Baldwin, and H.M. Krider. 1997. Effects of methionine on the cytoplasmic distribution of actin and tubulin during neural tube closure in rat embryos. *Proc. Natl. Acad. Sci. USA.* 94:543–548.

Najbauer, J., and D.W. Aswad. 1990. Diversity of methyl acceptor proteins in rat pheochromocytoma (PC12) cells revealed after treatment with adenosine dialdehyde. *J. Biol. Chem.* 265:12717–12721.

Parish, C.A., A.V. Smrcka, and R.R. Rando. 1995. Functional significance of βγ-subunit carboxyl methylation for the activation of phospholipase C and phosphoinositide 3-kinase. *Biochemistry.* 34:7722–7727.

Pawson, T. 1985. Protein modules and signaling networks. *Nature (Lond.).* 373: 573–580.

Peng, X., L.A. Greene, D.R. Kaplan, and R.M. Stephens. 1995. Deletion of a

conserved juxtamembrane sequence in trk abolishes NGF-promoted neuritogenesis. *Neuron.* 15:395–406.

Philips, M.R., M.H. Pillinger, R. Staud, C. Volker, M.G. Rosenfeld, G. Weissmann, and J.B. Stock. 1993. Carboxyl methylation of Ras-related proteins during signal transduction in neutrophils. *Science (Wash. DC).* 259:868–872.

Philips, M.R., R. Staud, M. Pillinger, A. Feoktistov, C. Volker, J.B. Stock, and G. Weissmann. 1995. Activation-dependent carboxyl methylation of neutrophil G-protein γ subunit. *Proc. Natl. Acad Sci. USA.* 92:2283–2287.

Qui, M.-S., and S.H. Green. 1992. PC12 cell neuronal differentiation is associated with prolonged $p21^{ras}$ activity and consequent prolonged ERK activity. *Neuron.* 9:705–717.

Rando, R.R. 1996. Chemical biology of protein isoprenylation/methylation. *Biochim. Biophys. Acta.* 1300:5–16.

Razin, A., and H. Cedar. 1991. DNA methylation and gene expression. *Microbiol. Rev.* 55:451–458.

Rukenstein, A., R.E. Rydel, and L.A. Greene. 1991. Multiple agents rescue PC12 cells from serum-free cell death by translation, and transcription-dependent mechanisms. *J. Neurosci.* 11:2557–2563.

Seeley, P.J., A. Rukenstein, J.L. Connolly, and L.A. Greene. 1984. Differential inhibition of nerve growth factor and epidermal growth factor on the PC12 pheochromocytoma line. *J. Cell Biol.* 98:417–426.

Shapiro, M.J., I. Chakrabarti, and D.E. Koshland. 1995. Contributions made by individual methylation sites of *Escherichia coli* aspartate receptor to chemotactic behavior. *Proc. Natl. Acad. Sci. USA.* 92:1053–1056.

Soto, A.M., and C. Sonnenschein. 1985. The role of estrogens on the proliferation of human breast tumor cells (MCF-7). *J. Steroid Biochem.* 23:87–94.

Tischler, A.S., and L.A. Greene. 1978. Morphologic and cytochemical properties of a clonal line of rat adrenal pheochromocytoma cells which respond to nerve growth factor. *Lab. Invest.* 39:77–89.

Turowski, P., A. Fernandez, B. Favre, N.J.C. Lamb, and B.A. Hemmings. 1995. Differential methylation and altered conformation of cytoplasmic and nuclear forms of protein phosphatase 2A during cell cycle progression. *J. Cell Biol.* 129:397–410.

Vojtek, A.B., and J.A. Cooper. 1995. Rho family members: activators of MAP kinase cascades. *Cell.* 82:527–529.

Xie, H., and S. Clarke. 1994. Protein phosphatase 2A is reversibly modified by methyl esterification of its C-terminal leucine residue in bovine brain. *J. Biol. Chem.* 269:1981–1984.